SPECIAL COLLECTOR'S EDITION

FALL 2020

SCIENTIFIC AMERICAN

TRUTH VS. LIES

IT'S NEVER BEEN MORE IMPORTANT TO UNDERSTAND THE SCIENCE OF MISINFORMATION AND DECEPTION AND HOW TO KNOW WHAT'S REAL

INSIDE

- Facebook and fake news are polarizing society
- Antiscience thinking
- Implicit bias affects our reality
- Battling disinformation
- The psychology behind conspiracy theories

© 2020 Scientific American

Scientific American Unlimited

Perfect for science fans, gain access to all of our publications, apps & the full website experience.

Digital archive access back to 1845 including articles by Einstein, Curie and other timeless luminaries in more than 7,000 issues!

12 print and digital issues of *Scientific American* per year

More than 150 eBooks and Collector's Editions

Access to *Scientific American Mind*, *Space & Physics* and *Health & Medicine*

More than 200 articles per month, including news and commentary, on ScientificAmerican.com

sciam.com/unlimited

Scientific American is a registered trademark of Springer Nature America, Inc

SCIENTIFIC AMERICAN®
ESTABLISHED 1845

Truth vs. Lies is published by the staff of Scientific American, with project management by:

Editor in Chief: **Laura Helmuth**
Managing Editor: **Curtis Brainard**
Senior Editor, Collections: **Andrea Gawrylewski**

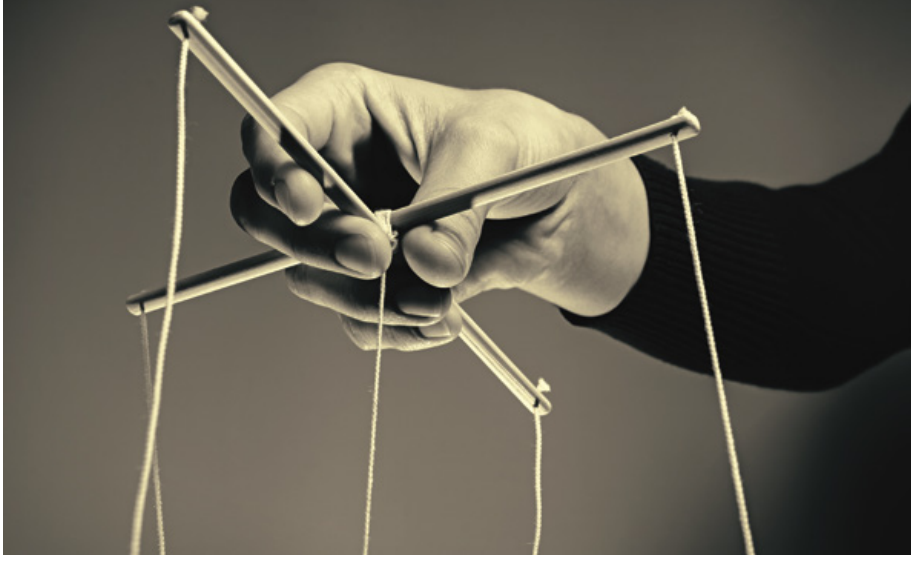
Creative Director: **Michael Mrak**
Issue Designer: **Lawrence R. Gendron**
Senior Graphics Editor: **Jen Christiansen**
Associate Graphics Editor: **Amanda Montañez**
Photography Editor: **Monica Bradley**
Assistant Photo Editor: **Liz Tormes**

Copy Director: **Maria-Christina Keller**
Senior Copy Editors: **Angelique Rondeau, Daniel C. Schlenoff, Aaron Shattuck**
Managing Production Editor: **Richard Hunt**
Prepress and Quality Manager: **Silvia De Santis**

Editorial Administrator: **Ericka Skirpan**
Executive Assistant Supervisor: **Maya Harty**

President: **Dean Sanderson**
Executive Vice President: **Michael Florek**
Vice President, Commercial: **Andrew Douglas**
Vice President, Magazines, Editorial and Publishing: **Stephen Pincock**
Publisher and Vice President: **Jeremy A. Abbate**
Associate Vice President, Business Development: **Diane McGarvey**

Marketing Director, Institutional Partnerships and Customer Development: **Jessica Cole**
Programmatic Product Manager: **Zoya Lysak**
Director, Integrated Media: **Matt Bondlow**
Senior Administrator, Executive Services: **May Jung**
Senior Marketing Manager: **Christopher Monello**
Product Manager: **Ian Kelly**
Senior Web Producer: **Jessica Ramirez**
Senior Commercial Operations Coordinator: **Christine Kaelin**
Custom Publishing Editor: **Lisa Pallatroni**
Rights and Permissions Manager: **Felicia Ruocco**
Head, Communications, USA: **Rachel Scheer**
Press Manager: **Sarah Hausman**
Production Controller: **Madelyn Keyes-Milch**
Advertising Production Controller: **Dan Chen**



GETTY IMAGES

FROM THE EDITOR

21st-Century Gaslighting

Untruths come in many shapes and sizes. All are antithetical to science, but the most flagrant variety—gaslighting, in which someone denies the reality that other people are experiencing—has become a greater threat than ever. Throughout a year of tragedy and strife, including the coronavirus pandemic and the Black Lives Matter protests, we've seen leaders and other public figures force people to doubt their experiences and thoughts. These manipulators don't just ignore data and evidence; they deny their very existence.

How do we navigate what philosopher Kathleen Higgins describes as the post-truth era, where there is no longer even an expectation that politicians or pundits will be honest (*page 70*)? This issue, we hope, will serve in some measure as an antidote to the poison of gaslighting and other forms of mendacity. Never has it been more important to understand the science of deception and what can be done to combat it.

For starters, human perception is inherently subjective (*page 4*). We may believe that we are thinking in an open-minded way, but most people latch onto ideas that reinforce their own preconceived beliefs—even if doing so prevents them from literally seeing new solutions (*page 12*). Such ingrained implicit bias has served us well in the course of evolution, but in the modern era it more often than not leads us astray (*page 18*).

Nowhere are our failings at objective reasoning more exploitable than on social media. Facebook and other platforms enable the spread of misinformation that sows social unrest—in particular, meme culture has been shown to propagate lies and increase polarization (*page 40*). We users are the unwitting vectors of this threat (*page 39*). (Watch our short documentary on the making of a deepfake video at https://bit.ly/3itSiek for a close look into another way that technology enables deceit.)

Malicious foreign agents invade those same platforms, and hackers could go further and tamper with vulnerable voting systems in this country, endangering the very processes of democracy (*page 62*). Civic life suffers because of these malevolent forces. Turmoil, anxiety and a sense that society is in jeopardy make people susceptible to conspiracy theories (*page 74*) and the kind of polarization where winning an argument is more important than understanding opponents' viewpoints (*pages 66* and *72*).

Although the human mind comes equipped with built-in obstacles to objective thinking, it doesn't mean we have to give in to ignorance and bias. Douglas T. Kenrick and his co-authors offer simple interventions that can make us more open-minded, scientific thinkers (*page 84*). In fact, scientists can look to philosophy to aid in some self-examination about how much, in the hands of subjective creatures, the tools of science can ultimately discover (*page 110*).

The common theme in many of these seemingly abysmal examinations of the state of our societal affairs is a heartening bright spot. By just being aware of how we perceive information, we can protect ourselves from disinformation and gaslighting. We don't have to agree, but at least we'll be anchored in what is real and what is not.

Andrea Gawrylewski
Senior Editor, Collections
editors@sciam.com

© 2020 Scientific American

SPECIAL EDITION

SCIENTIFIC AMERICAN

Volume 29, Number 4, Fall 2020

GRAPPLING WITH REALITY

4 Our Inner Universes
Reality is constructed by the brain, and no two brains are exactly alike. *By Anil K. Seth*

12 Why Good Thoughts Block Better Ones
While we are woking through a problem, the brain's tendency to stick with familiar ideas can cause us to ignore a superior solution. *By Merim Bilalić and Peter McLeod*

18 How to Think about "Implicit Bias"
Amidst a controversy, it's important to remember that implicit bias is real—and it matters. *By Keith Payne, Laura Niemi, John M. Doris*

20 Why We Trust Lies
The most effective misinformation starts with seeds of truth. *By Cailin O'Connor and James Owen Weatherall*

DECISION-MAKING

26 Tough Calls
How we make decisions in the face of incomplete knowledge and uncertainty. *By Baruch Fischhoff*

32 Confronting Unknowns
Interpreting uncertainty in common forms of data visualizations. *By Jessica Hullman*

36 Which Experts Should You Listen to during the Pandemic?
It should be obvious, but too many people don't realize they should follow those who have actual expertise. *By Nathan Ballantyne and David Dunning*

38 Big Data and Small Decisions
A deluge of facts can be a problem. *By Zeynep Tufekci*

TECHNOLOGY'S INFLUENCE

39 When "Like" Is a Weapon
Everyone is an agent in the new information warfare. *By the Editors*

40 A New World Disorder
Our willingness to share content without thinking is exploited to spread disinformation. *By Claire Wardle*

46 Inside the Echo Chamber
Data scientists are studying how information spreading online influences our social dynamics and what, if anything, can be done to smooth polarization. *By Walter Quattrociocchi*

50 The Shared Past That Wasn't
How Facebook, fake news and friends are altering memories and changing history. *By Laura Spinney*

56 Clicks, Lies and Videotape
Artificial intelligence is making it possible for anyone to manipulate audio and video. The biggest threat is that we stop trusting anything at all. *By Brooke Borel*

62 How to Defraud Democracy
A worst-case cyberwarfare scenario for the 2020 American presidential election. *By J. Alex Halderman as told to Jen Schwartz*

© 2020 Scientific American

POLITICS

66 The Tribalism of Truth
As political polarization grows, the arguments we have with one another may be shifting our understanding of truth itself.
By Matthew Fisher, Joshua Knobe, Brent Strickland and Frank C. Keil

70 Post-Truth: A Guide for the Perplexed
If politicians can lie without condemnation, what are scientists to do? *By Kathleen Higgins*

72 Why People "Fly from Facts"
Research shows the appeal of untestable beliefs and how they lead to a polarized society. *By Troy Campbell and Justin Friesen*

74 Why We Believe Conspiracy Theories
Baseless theories threaten our safety and democracy. It turns out that specific emotions make people prone to such thinking.
By Melinda Wenner Moyer

80 Contagious Dishonesty
Dishonesty begets dishonesty, rapidly spreading unethical behavior through a society.
By Dan Ariely and Ximena Garcia-Rada

FOUNDATIONS OF SCIENCE

84 The Science of Antiscience Thinking
Convincing people who doubt the validity of climate change and evolution to change their beliefs requires overcoming a set of ingrained cognitive biases.
By Douglas T. Kenrick, Adam B. Cohen, Steven L. Neuberg and Robert B. Cialdini

90 How Professional Truth Seekers Search for Answers
Nine experts describe how they sort signal from noise.
As told to Brooke Borel

96 The Truth about Scientific Models
They don't necessarily try to predict what will happen—but they can help us understand possible futures. *By Sabine Hossenfelder*

98 A Significant Problem
Standard scientific methods are under fire. Will anything change?
By Lydia Denworth

104 The Roots of Science Denial
It has nothing do with science itself. *By Katharine Hayhoe as told to Jen Schwartz*

108 Message Control
Scientists are trying new ways to win over a skeptical public.
By Brooke Borel

110 How Much Can We Know?
The reach of the scientific method is constrained by the limitations of our tools and the intrinsic impenetrability of some of nature's deepest questions.
By Marcelo Gleiser

DEPARTMENTS

FROM THE EDITOR
1 21st-Century Gaslighting

END NOTE
112 Ya Know?
Beyond the unknown unknowns is what's unknowable.
By Steve Mirsky

Articles in this special issue are updated or adapted from previous issues of *Scientific American* and *Nature* and from ScientificAmerican.com.

Copyright © 2020 Scientific American, a division of Springer Nature America, Inc. All rights reserved.

Scientific American Special (ISSN 1936-1513), Volume 29, Number 4, Fall 2020, published by Scientific American, a division of Springer Nature America, Inc., 1 New York Plaza, Suite 4600, New York, N.Y. 10004-1562. Canadian BN No. 127387652RT; TVQ1218059275 TQ0001. To purchase additional quantities: U.S., $13.95 each; elsewhere, $17.95 each. Send payment to Scientific American Back Issues, P.O. Box 3187, Harlan IA 51537.
Inquiries: fax 212-355-0408 or telephone 212-451-8415. Printed in U.S.A.

© 2020 Scientific American

OUR INNER UNIVERSES

Reality is constructed by the brain, and no two brains are exactly alike

By Anil K. Seth

"We do not see things as they are, we see them as we are."
—from *Seduction of the Minotaur,* by Anaïs Nin (1961)

Illustration by Brook VanDevelder

© 2020 Scientific American

© 2020 Scientific American

ON THE 10TH OF APRIL 2019 POPE FRANCIS, PRESIDENT SALVA KIIR of South Sudan and former rebel leader Riek Machar sat down together for dinner at the Vatican. They ate in silence, the start of a two-day retreat aimed at reconciliation from a civil war that has killed some 400,000 people since 2013. At about the same time in my laboratory at the University of Sussex in England, Ph.D. student Alberto Mariola was putting the finishing touches on a new experiment in which volunteers experience being in a room that they believe is there but that is not. In psychiatry clinics across the globe, people arrive complaining that things no longer seem "real" to them, whether it is the world around them or their own selves. In the fractured societies in which we live, what is real—and what is not—seems to be increasingly up for grabs. Warring sides may experience and believe in different realities. Perhaps eating together in silence can help because it offers a small slice of reality that can be agreed on, a stable platform on which to build further understanding.

IN BRIEF

The reality we perceive is not a direct reflection of the external objective world. **Instead it is the product** of the brain's predictions about the causes of incoming sensory signals. **The property of realness** that accompanies our perceptions may serve to guide our behavior so that we respond appropriately to the sources of sensory signals.

We need not look to war and psychosis to find radically different inner universes. In 2015 a badly exposed photograph of a dress tore across the Internet, dividing the world into those who saw it as blue and black (me included) and those who saw it as white and gold (half my lab). Those who saw it one way were so convinced they were right—that the dress truly was blue and black or white and gold—that they found it almost impossible to believe that others might perceive it differently.

We all know that our perceptual systems are easy to fool. The popularity of visual illusions is testament to this phenomenon. Things seem to be one way, and they are revealed to be another: two lines appear to be different lengths, but when measured they are exactly the same; we see movement in an image we know to be still. The story usually told about illusions is that they exploit quirks in the circuitry of perception, so that what we perceive deviates from what is there. Implicit in this story, however, is the assumption that a properly functioning perceptual system will render to our consciousness things precisely as they are.

The deeper truth is that perception is never a direct window onto an objective reality. All our perceptions are active constructions, brain-based best guesses at the nature of a world that is forever obscured behind a sensory veil. Visual illusions are fractures in the Matrix, fleeting glimpses into this deeper truth.

Take, for example, the experience of color—say, the bright red of the coffee mug on my desk. The mug really does seem to be red: its redness seems as real as its roundness and its solidity. These features of my experience seem to be truly existent properties of the world, detected by our senses and revealed to our mind through the complex mechanisms of perception.

Yet we have known since Isaac Newton that colors do not exist out there in the world. Instead they are cooked up by the brain from mixtures of different wavelengths of colorless electromagnetic radiation. Colors are a clever trick that evolution has hit on to help the brain keep track of surfaces under changing lighting conditions. And we humans can sense only a tiny slice of the full electromagnetic spectrum, nestled between the lows of infrared and the highs of ultraviolet. Every color we perceive, every part of the totality of each of our visual worlds, comes from this thin slice of reality.

Just knowing this is enough to tell us that perceptual experience cannot be a comprehensive representation of an external objective world. It is both less than that and more than that. The reality we experience—the way things *seem*—is not a direct reflection of what is actually out there. It is a clever construction by the brain, for the brain. And if my brain is different from your brain, my reality may be different from yours, too.

THE PREDICTIVE BRAIN

IN PLATO'S *Allegory of the Cave,* prisoners are chained to a blank wall all their lives, so that they see only the play of shadows cast by objects passing by a fire behind them, and they give the shadows names because for them the shadows are what is real. A thousand years

© 2020 Scientific American

later, but still a thousand years ago, Arabian scholar Ibn al-Haytham wrote that perception, in the here and now, depends on processes of "judgment and inference" rather than involving direct access to an objective reality. Hundreds of years later again Immanuel Kant realized that the chaos of unrestricted sensory data would always remain meaningless without being given structure by preexisting conceptions or "beliefs," which for him included a priori frameworks such as space and time. Kant's term "noumenon" refers to a "thing in itself"—*Ding an sich*—an objective reality that will always be inaccessible to human perception.

Today these ideas have gained a new momentum through an influential collection of theories that turn on the idea that the brain is a kind of prediction machine and that perception of the world—and of the self within it—is a process of brain-based prediction about the causes of sensory signals.

These new theories are usually traced to German physicist and physiologist Hermann von Helmholtz, who in the late 19th century proposed that perception is a process of unconscious inference. Toward the end of the 20th century Helmholtz's notion was taken up by cognitive scientists and artificial-intelligence researchers, who reformulated it in terms of what is now generally known as predictive coding or predictive processing.

The central idea of predictive perception is that the brain is attempting to figure out what is out there in the world (or in here, in the body) by continually making and updating best guesses about the causes of its sensory inputs. It forms these best guesses by combining prior expectations or "beliefs" about the world, together with incoming sensory data, in a way that takes into account how reliable the sensory signals are. Scientists usually conceive of this process as a form of Bayesian inference, a framework that specifies how to update beliefs or best guesses with new data when both are laden with uncertainty.

In theories of predictive perception, the brain approximates this kind of Bayesian inference by continually generating predictions about sensory signals and comparing these predictions with the sensory signals that arrive at the eyes and the ears (and the nose and the fingertips, and all the other sensory surfaces on the outside and inside of the body). The differences between predicted and actual sensory signals give rise to so-called prediction errors, which are used by the brain to update its predictions, readying it for the next round of sensory inputs. By striving to minimize sensory-prediction errors everywhere and all the time, the brain implements approximate Bayesian inference, and the resulting Bayesian best guess is what we perceive.

To understand how dramatically this perspective shifts our intuitions about the neurological basis of perception, it is helpful to think in terms of bottom-up and top-down directions of signal flow in the brain. If we assume that perception is a direct window onto an external reality, then it is natural to think that the content of perception is carried by bottom-up signals—those that flow from the sensory surfaces inward. Top-down signals might contextualize or finesse what is perceived, but nothing more. Call this the "how things seem" view because it seems as if the world is revealing itself to us directly through our senses.

POORLY EXPOSED photograph of a dress appears blue and black to some people, white and gold to others.

The prediction machine scenario is very different. Here the heavy lifting of perception is performed by the top-down signals that convey perceptual predictions, with the bottom-up sensory flow serving only to calibrate these predictions, keeping them yoked, in some appropriate way, to their causes in the world. In this view, our perceptions come from the inside out just as much as, if not more than, from the outside in. Rather than being a passive registration of an external objective reality, perception emerges as a process of active construction—a controlled hallucination, as it has come to be known.

Why controlled hallucination? People tend to think of hallucination as a kind of false perception, in clear contrast to veridical, true-to-reality, normal perception. The prediction machine view suggests instead a continuity between hallucination and normal perception.

SWIKED.TUMBLR.COM

© 2020 Scientific American

Origins of Perception

The classical view of perception (*blue panel*) holds that it is a direct window onto an external reality. Sensory signals flow from the bottom up, entering the brain through receptors in our eyes, ears, nose, tongue and skin to reveal the outside world to us as it is. Top-down signals within the brain serve only to finesse what is perceived. In the prediction machine view of perception (*green panel*), in contrast, perceptual content is carried by top-down predictions made by the brain based on prior experience. Bottom-up signals **function mainly to convey prediction errors, which rein in the brain's hypotheses. Perception is thus a controlled hallucination in this model.**

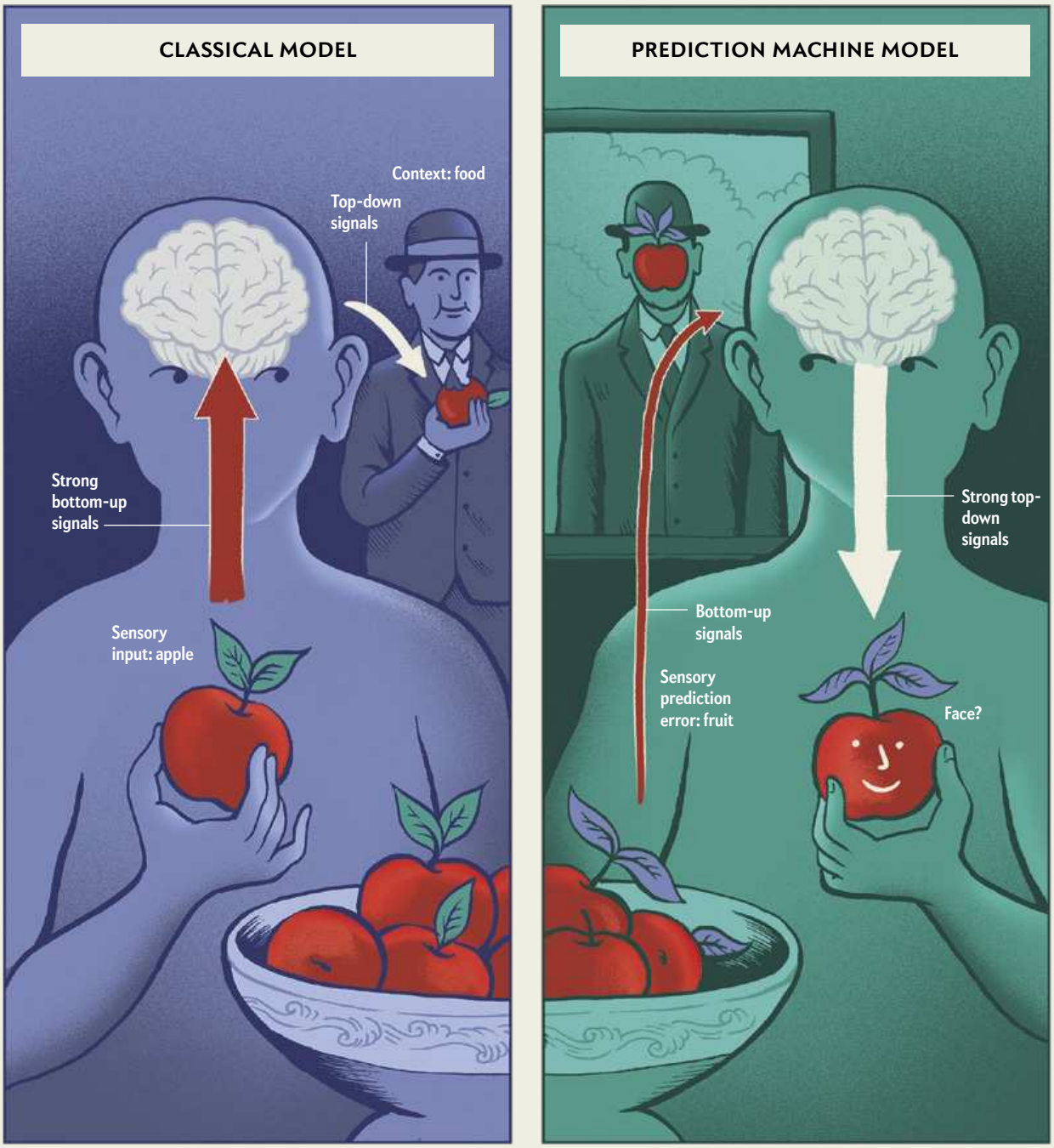
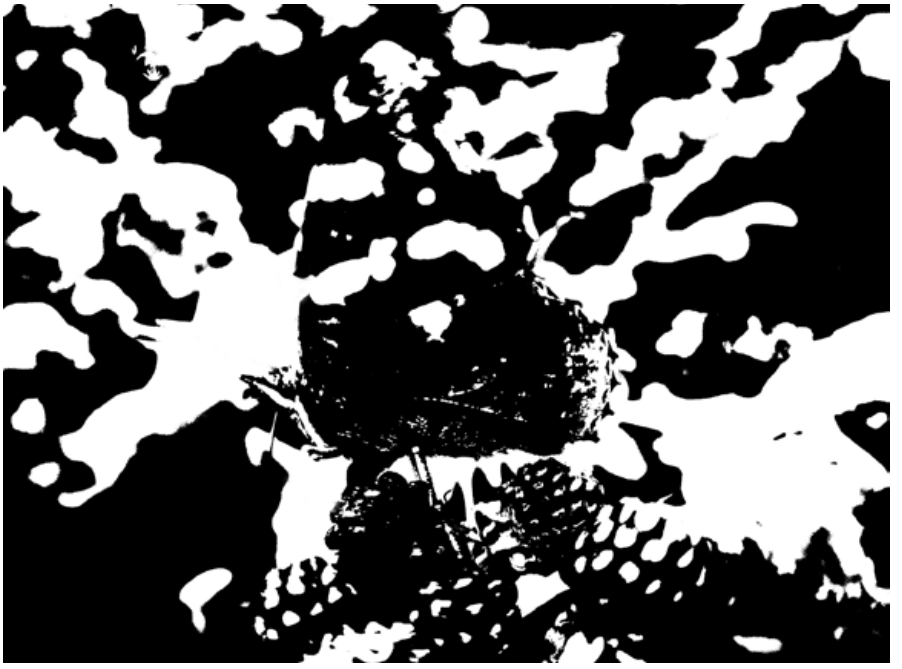


Illustration by Matteo Farinella

© 2020 Scientific American

Both depend on an interaction between top-down, brain-based predictions and bottom-up sensory data, but during hallucinations, sensory signals no longer keep these top-down predictions appropriately tied to their causes in the world. What we call hallucination, then, is just a form of uncontrolled perception, just as normal perception is a controlled form of hallucination.

This view of perception does not mean that nothing is real. Writing in the 17th century, English philosopher John Locke made an influential distinction between "primary" and "secondary" qualities. Primary qualities of an object, such as solidity and occupancy of space, exist independently of a perceiver. Secondary qualities, in contrast, exist only in relation to a perceiver—color is a good example. This distinction explains why conceiving of perception as controlled hallucination does not mean it is okay to jump in front of a bus. This bus has primary qualities of solidity and space occupancy that exist independently of our perceptual machinery and that can do us injury. It is the way in which the bus appears to us that is a controlled hallucination, not the bus itself.



TWO-TONE IMAGE looks like a mess of black-and-white splotches, until you see the full image on page 11.

RICHARD ARMSTRONG *Getty Images*

TRIPPING IN THE LAB

A GROWING BODY of evidence supports the idea that perception is controlled hallucination, at least in its broad outlines. A 2015 study by Christoph Teufel of Cardiff University in Wales and his colleagues offers a striking example. In this study, patients with early-stage psychosis who were prone to hallucinations were compared with healthy individuals on their ability to recognize so-called two-tone images.

Take a look at the photograph above—a sample of a two-tone image. Probably all you will see is a bunch of black-and-white splotches. Now, after you have read the rest of this sentence, look at the image on page 11. Then have another look at the first photo; it ought to look rather different. Where previously there was a splotchy mess, there are now distinct objects, and something is happening.

What I find remarkable about this exercise is that in your second examination of the photo above, the sensory signals arriving at your eyes have not changed at all from the first time you saw it. All that has changed are your brain's predictions about the causes of these sensory signals. You have acquired a new high-level perceptual expectation, and this is what changes what you consciously see.

If you show people many of these two-tone images, each followed by the full picture, they might subsequently be able to identify a good proportion of two-tone images, though not all of them. In Teufel's study, people with early-stage psychosis were better at recognizing two-tone images after having seen the full image than were healthy control subjects. In other words, being hallucination-prone went along with perceptual priors having a stronger effect on perception. This is exactly what would be expected if hallucinations in psychosis depended on an overweighting of perceptual priors so that they overwhelmed sensory prediction errors, unmooring perceptual best guesses from their causes in the world.

Recent research has revealed more of this story. Phil Corlett of Yale University and his colleagues paired lights and sounds in a simple design to engender expectations among their study subjects of whether or not a light would appear on a given experimental trial. They combined this design with brain imaging to uncover some of the brain regions implicated in predictive perception. When they looked at the data, Corlett and his team were able to identify regions such as the superior temporal sulcus, deep in the temporal lobe of the cortex, that were specifically associated with top-down predictions about auditory sensations. This is an exciting new development in mapping the brain basis of controlled hallucinations.

In my lab we have taken a different approach to exploring the nature of perception and hallucination. Rather than looking into the brain directly, we decided to simulate the influence of overactive perceptual priors using a unique virtual-reality setup masterminded by our resident VR guru, Keisuke Suzuki. We call it, with tongue firmly in cheek, the "hallucination machine."

Using a 360-degree camera, we first recorded panoramic video footage of a busy square in the University of Sussex campus on a Tuesday at lunchtime. We then

© 2020 Scientific American

processed the footage through an algorithm based on Google's AI program DeepDream to generate a simulated hallucination. What happens is that the algorithm takes a so-called neural network—one of the workhorses of AI—and runs it backward. The network we used had been trained to recognize objects in images, so if you run it backward, updating the network's input instead of its output, the network effectively projects what it "thinks" is there onto and into the image. Its predictions overwhelm the sensory inputs, tipping the balance of perceptual best guessing toward these predictions. Our particular network was good at classifying different breeds of dogs, so the video became unusually suffused by dog presences.

Many people who have viewed the processed footage through the VR headset have commented that the experience is rather reminiscent not of the hallucinations of psychosis but of the exuberant phenomenology of psychedelic trips.

By implementing the hallucination machine in slightly different ways, we could generate different kinds of conscious experience. For example, running the neural network backward from one of its middle layers, rather than from the output layer, leads to hallucinations of object parts, rather than whole objects. As we look ahead, this method will help us match specific features of the computational architecture of predictive perception to specific aspects of what experiences of hallucinations are like. And by understanding hallucinations better, we will be able to understand normal experience better, too, because predictive perception is at the root of all our perceptual experience.

OUR PERCEPTIONS COME FROM THE INSIDE OUT JUST AS MUCH AS, IF NOT MORE THAN, FROM THE OUTSIDE IN.

THE PERCEPTION OF REALITY

ALTHOUGH THE HALLUCINATION MACHINE is undoubtedly trippy, people who experience it are fully aware that what they are experiencing is not real. Indeed, despite rapid advances in VR technology and computer graphics, no current VR setup delivers an experience that is sufficiently convincing to be indistinguishable from reality.

This is the challenge we took up when designing a new "substitutional reality" setup at Sussex—the one we were working on when Pope Francis convened the retreat with Salva Kiir and Riek Machar. Our aim was to create a system in which volunteers would experience an environment as being real—and believe it to be real—when in fact it was not real.

The basic idea is simple. We again prerecorded some panoramic video footage, this time of the interior of our VR lab rather than of an outside campus scene. People coming to the lab are invited to sit on a stool in the middle of the room and to put on a VR headset that has a camera attached to the front. They are encouraged to look around the room and to see the room as it actually is, via the camera. But at some point, without telling them, we switch the feed so that the headset now displays not the live real-world scene but rather the prerecorded panoramic video. Most people in this situation continue to experience what they are seeing as real even though it is now a fake prerecording. (This is actually very tricky to pull off in practice—it requires careful color balancing and alignment to avoid people noticing any difference that would tip them off to the shift.)

I find this result fascinating because it shows that it is possible to have people experience an unreal environment as being fully real. This demonstration alone opens new frontiers for VR research: we can test the limits of what people will experience, and believe, to be real. It also allows us to investigate how experiencing things as being real can affect other aspects of perception. Right now we are running an experiment to find out whether people are worse at detecting unexpected changes in the room when they believe that what they are experiencing is real. If things do turn out this way (the study is ongoing), that finding would support the idea that the perception of things as being real itself acts as a high-level prior that can substantively shape our perceptual best guesses, affecting the contents of what we perceive.

THE REALITY OF REALITY

THE IDEA THAT THE WORLD of our experience might not be real is an enduring trope of philosophy and science fiction, as well as of late-night pub discussions. Neo in *The Matrix* takes the red pill, and Morpheus shows him how what he thought was real is an elaborate simulation, while the real Neo lies prone in a human body farm, a brain-in-a-vat power source for a dystopian AI. Philosopher Nick Bostrom of the University of Oxford has famously argued, based largely on statistics, that we are likely to be living inside a computer simulation created in a posthuman age. I disagree with this argument because it assumes that consciousness can be simulated—I do not think that this is a safe assumption—but it is thought-provoking nonetheless.

Although these chunky metaphysical topics are fun to chew on, they are probably impossible to resolve. Instead what we have been exploring throughout this article is the relation between appearance and reality in our conscious perceptions, where part of this appearance is the appearance of being real itself.

The central idea here is that perception is a process of active interpretation geared toward adaptive inter-

© 2020 Scientific American

action with the world through the body rather than a recreation of the world within the mind. The contents of our perceptual worlds are controlled hallucinations, brain-based best guesses about the ultimately unknowable causes of sensory signals. And for most of us, most of the time, these controlled hallucinations are experienced as real. As Canadian rapper and science communicator Baba Brinkman suggested to me, when we agree about our hallucinations, maybe that is what we call reality.

But we do not always agree, and we do not always experience things as real. People with dissociative psychiatric conditions such as derealization or depersonalization syndrome report that their perceptual worlds, even their own selves, lack a sense of reality. Some varieties of hallucination, various psychedelic hallucinations among them, combine a sense of unreality with perceptual vividness, as does lucid dreaming. People with synesthesia consistently have additional sensory experiences, such as perceiving colors when viewing black letters, which they recognize as not real. Even with normal perception, if you look directly at the sun you will experience the subsequent retinal afterimage as not being real. There are many such ways in which we experience our perceptions as not fully real.

What this means to me is that the property of realness that attends most of our perceptions should not be taken for granted. It is another aspect of the way our brain settles on its Bayesian best guesses about its sensory causes. One might therefore ask what purpose it serves. Perhaps the answer is that a perceptual best guess that includes the property of being real is usually more fit for purpose—that is, better able to guide behavior—than one that does not. We will behave more appropriately with respect to a coffee cup, an approaching bus or our partner's mental state when we experience it as really existing.

But there is a trade-off. As illustrated by the dress illusion, when we experience things as being real, we are less able to appreciate that our perceptual worlds may differ from those of others. (One leading explanation for the differing perceptions of the garment holds that people who spend most of their waking hours in daylight see it as white and gold; night owls, who are mainly exposed to artificial light, see it as blue and black.) And even if these differences start out small, they can become entrenched and reinforced as we proceed to harvest information differently, selecting sensory data that are best aligned with our individual emerging models of the world and then updating our perceptual models based on these biased data. We are all familiar with this process from the echo chambers of social media and the newspapers we choose to read. I am suggesting that the same principles apply also at a deeper level, underneath our sociopolitical beliefs, right down to the fabric of our perceptual realities. They may even apply to our perception of being a self—the experience of being me or of being you—because the experience of being a self is itself a perception.

This is why understanding the constructive, creative mechanisms of perception has an unexpected social relevance. Perhaps once we can better appreciate the diversity of experienced realities scattered among the billions of perceiving brains on this planet, we will find new platforms on which to build a shared understanding and a better future—whether between sides in a civil war, followers of different political parties, or two people sharing a house and faced with washing the dishes. SA

RICHARD ARMSTRONG Getty Images

PERCEPTUAL SHIFT: Viewing this photograph changes what one consciously sees in the two-tone image on page 9.

Anil K. Seth is a professor of cognitive and computational neuroscience at the University of Sussex in England and co-director of the university's Sackler Center for Consciousness Science. His research focuses on the biological basis of consciousness.

MORE TO EXPLORE

Shift toward Prior Knowledge Confers a Perceptual Advantage in Early Psychosis and Psychosis-Prone Healthy Individuals. Christoph Teufel et al. in *Proceedings of the National Academy of Sciences USA,* Vol. 112, No. 43, pages 13,401–13,406; October 27, 2015.

A Deep-Dream Virtual Reality Platform for Studying Altered Perceptual Phenomenology. Keisuke Suzuki et al. in *Scientific Reports,* Vol. 7, Article No. 15982; November 22, 2017.

Being a Beast Machine: The Somatic Basis of Selfhood. Anil K. Seth and Manos Tsakiris in *Trends in Cognitive Sciences,* Vol. 22, No. 11, pages 969–981; November 1, 2018.

scientificamerican.com/magazine/sa

© 2020 Scientific American

© 2020 Scientific American

WHY GOOD THOUGHTS BLOCK BETTER ONES

While we are working through a problem, the brain's tendency to stick with familiar ideas can literally blind us to superior solutions

By Merim Bilalić and Peter McLeod

Illustration by Danny Schwartz

© 2020 Scientific American

IN A CLASSIC 1942 EXPERIMENT, AMERICAN PSYCHOLOGIST ABRAHAM LUCHINS asked volunteers to do some basic math by picturing water jugs in their mind. Given three empty containers, for example, each with a different capacity—21, 127 and three units of water—the participants had to figure out how to transfer liquid between the containers to measure out precisely 100 units. They could fill and empty each jug as many times as they wanted, but they had to fill the vessels to their limits. The solution was to first fill the second jug to its capacity of 127 units, then empty it into the first to remove 21 units, leaving 106, and finally to fill the third jug twice to subtract six units for a remainder of 100. Luchins presented his volunteers with several more problems that could be solved with essentially the same three steps; they made quick work of them. Yet when he gave them a problem with a simpler and faster solution than the previous tasks, they failed to see it.

This time Luchins asked the participants to measure out 20 units of water using containers that could hold 23, 49 and three liquid units. The solution is obvious, right? Simply fill the first jug and empty it into the third one: 23 – 3 = 20. Yet many people in Luchins's experiment persisted to solve the easier problem the old way, emptying the second container into the first and then into the third twice: 49 – 23 – 3 – 3 = 20. And when Luchins gave them a problem that had a two-step solution—but could not be solved using the three-step method to which the volunteers had become accustomed—they gave up, saying it was impossible.

The water jug experiment is one of the most famous examples of the *Einstellung* effect: the human brain's dogged tendency to stick with a familiar solution to a problem—the one that first comes to mind—and to ignore alternatives. Often this type of thinking is a useful heuristic. Once you have hit on a successful method to, say, peel garlic, there is no point in trying an array of different techniques every time you need a new clove. The trouble with this cognitive shortcut, however, is that it sometimes blinds people to more efficient or appropriate solutions than the ones they already know.

Building on Luchins's early work, psychologists replicated the *Einstellung* effect in many different laboratory studies with both novices and experts exercising a range of mental abilities, but exactly how and why it happened was never clear. Recently, by recording the eye movements of highly skilled chess players, we have solved the mystery. It turns out that people under the influence of this cognitive shortcut are literally blind to certain details in their environment that could provide them with a more effective solution. New research also suggests that many different cognitive biases discovered by psychologists over the years—those in the courtroom and the hospital, for instance—are in fact variations of the *Einstellung* effect.

BACK TO SQUARE ONE

SINCE AT LEAST THE EARLY 1990S, psychologists have studied the *Einstellung* effect by recruiting chess players of varying skill levels, from amateur to grand master. In such experiments, researchers have presented players with specific arrangements of chess pieces on virtual chessboards and asked them to achieve a checkmate in as few moves as possible. Our own studies, for instance, provided expert chess players with scenarios in which they could accomplish a checkmate using a well-known sequence called smothered mate. In this five-step maneuver, the queen is sacrificed to draw one of the opponent's pieces onto a square to block off the king's escape route. The players also had the option to checkmate the king in just three moves with a much less familiar sequence. As in Luchins's water jug studies, most of the players failed to find the more efficient solution.

During some of these studies, we asked the players what was going through their mind. They said they had found the smothered mate solution and insisted they were searching for a shorter one, to no avail. But the verbal reports offered no insight into why they

IN BRIEF

The *Einstellung* effect is the brain's tendency to stick with the most familiar solution to a problem and stubbornly ignore alternatives.

Psychologists have known about this mental phenomenon since the 1940s, but only now do they have a solid understanding of how it happens.

In recent eye-tracking experiments, familiar ideas blinded chess players to areas of a chessboard that would have provided clues to better solutions.

© 2020 Scientific American

could not find the swifter solution. In 2007 we decided to try something a little more objective: tracking eye movements with an infrared camera. Which part of the board people looked at and how long they looked at different areas would unequivocally tell us which aspects of the problem they were noticing and ignoring.

In this experiment, we followed the gaze of five expert chess players as they examined a board that could be solved either with the longer smothered mate maneuver or with the shorter three-move sequence. After an average of 37 seconds, all the players insisted that the smothered mate was the speediest possible way to corner the king. When we presented them with a board that could be solved only with the three-sequence move, however, they found it with no problem. And when we told the players that this same swift checkmate had been possible in the previous chessboard, they were shocked. "No, it is impossible," one player exclaimed. "It is a different problem; it must be. I would have noticed such a simple solution." Clearly, the mere possibility of the smothered mate move was stubbornly masking alternative solutions. In fact, the *Einstellung* effect was powerful enough to temporarily lower expert chess masters to the level of much weaker players.

ANY DATA THAT DO NOT FIT THE SOLUTION OR THEORY WE HAVE ALREADY CLUNG TO ARE IGNORED OR DISCARDED.

The infrared camera revealed that even when the players said they were looking for a faster solution—and indeed believed they were doing so—they did not actually shift their gaze away from the squares they had already identified as part of the smothered mate move. In contrast, when presented with the one-solution chessboard, players initially looked at the squares and pieces important for the smothered mate and, once they realized it would not work, directed their attention toward other squares and soon hit on the shorter solution.

BASIS FOR BIAS

IN 2013 Heather Sheridan, now at the University of Albany, S.U.N.Y., and Eyal M. Reingold of the University of Toronto published studies that corroborate and complement our eye-tracking experiments. They presented 17 novice and 17 expert chess players with two different situations. In one scenario, a familiar checkmate maneuver such as the smothered mate was advantageous but second best to a distinct and less obvious solution. In the second situation, the more familiar sequence would be a clear blunder. As in our experiments, once amateurs and master chess players locked onto the helpful familiar maneuver, their eyes rarely drifted to squares that would clue them in to the better solution. When the well-known sequence was obviously a mistake, however, all the experts (and most of the novices) detected the alternative.

The *Einstellung* effect is by no means limited to controlled experiments in the lab or even to mentally challenging games like chess. Rather it is the basis for many cognitive biases. English philosopher, scientist and essayist Francis Bacon was especially eloquent about one of the most common forms of cognitive bias in his 1620 book *Novum Organum:* "The human understanding when it has once adopted an opinion … draws all things else to support and agree with it. And though there be a greater number and weight of instances to be found on the other side, yet these it either neglects or despises, or else by some distinction sets aside and rejects.… Men … mark the events where they are fulfilled, but where they fail, though this happen much oftener, neglect and pass them by. But with far more subtlety does this mischief insinuate itself into philosophy and the sciences, in which the first conclusion colours and brings into conformity with itself all that comes after."

In the 1960s English psychologist Peter Wason gave this particular bias a name: "confirmation bias." In controlled experiments, he demonstrated that even when people attempt to test theories in an objective way, they tend to seek evidence that confirms their ideas and to ignore anything that contradicts them.

In *The Mismeasure of Man,* for example, Stephen Jay Gould of Harvard University reanalyzed data cited by researchers trying to estimate the relative intelligence of different racial groups, social classes and sexes by measuring the volumes of their skulls or weighing their brains, on the assumption that intelligence was correlated with brain size. Gould uncovered massive data distortion. On discovering that French brains were on average smaller than their German counterparts, French neurologist Paul Broca explained away the discrepancy as a result of the difference in average body size between citizens of the two nations. After all, he could not accept that the French were less intelligent than the Germans. Yet when he found that women's brains were smaller than those in men's noggins, he did not apply the same correction for body size, because he did not have any problem with the idea that women were less intelligent than men.

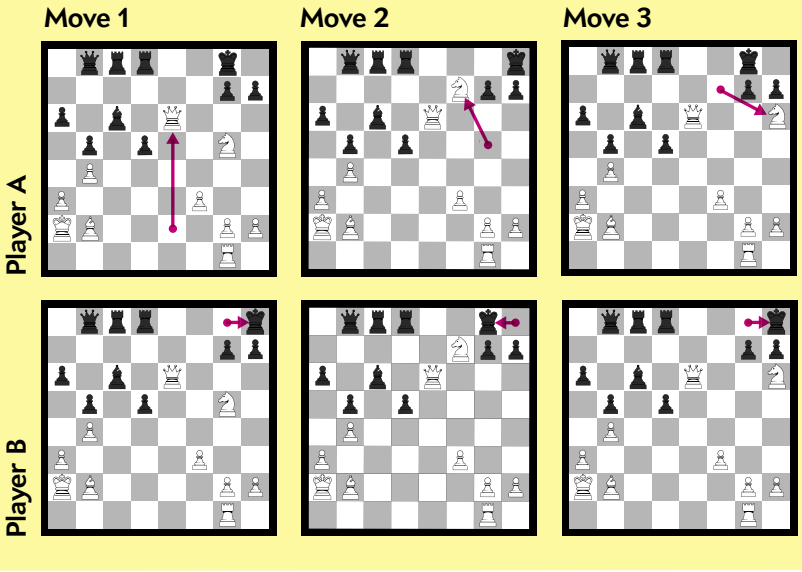
Somewhat surprisingly, Gould concluded that Broca and others like him were not as reprehensible as we might think. "In most cases discussed in this book we can be fairly certain that biases … were unknowingly influential and that scientists believed they were pursuing unsullied truth," Gould wrote. In other words, just as we observed in our chess experiments, comfortably familiar ideas blinded Broca and his contemporaries to the errors in their reasoning. Here is the real danger of the *Einstellung* effect. We may believe that we are thinking in an open-minded way, completely unaware that our brain is selectively directing attention away from aspects of our environment that could inspire new thoughts. Any data that do not fit the solution or theory we have already clung to are ignored or discarded.

The surreptitious nature of confirmation bias has unfortunate consequences in everyday life, as documented in studies on decision-making among doctors and juries. In a review of errors in medical thought, physician Jerome Groopman noted that in most cases of misdiagnosis, "the doctors didn't stumble because of their ignorance of clinical facts; rather, they missed diagnoses because they fell into cognitive traps." When doctors inherit a patient from another doctor, for example, the first clinician's diagnosis can blind the second to important and contradictory details of the patient's health that might change the diagnosis. It is easier to just accept the diagnosis—the "solution"—that is already in front of them than to rethink the entire situation. Similarly, radiologists examining chest x-rays often fixate on the first abnormality they find and fail to notice further signs of illness

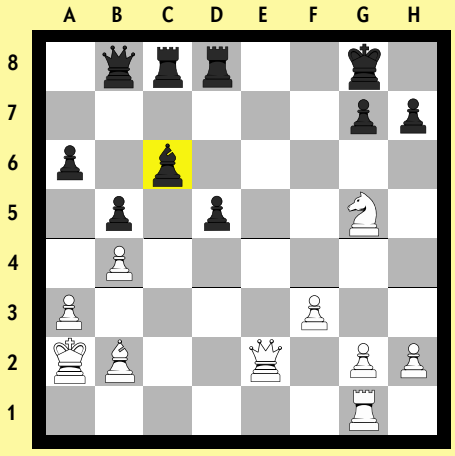
© 2020 Scientific American

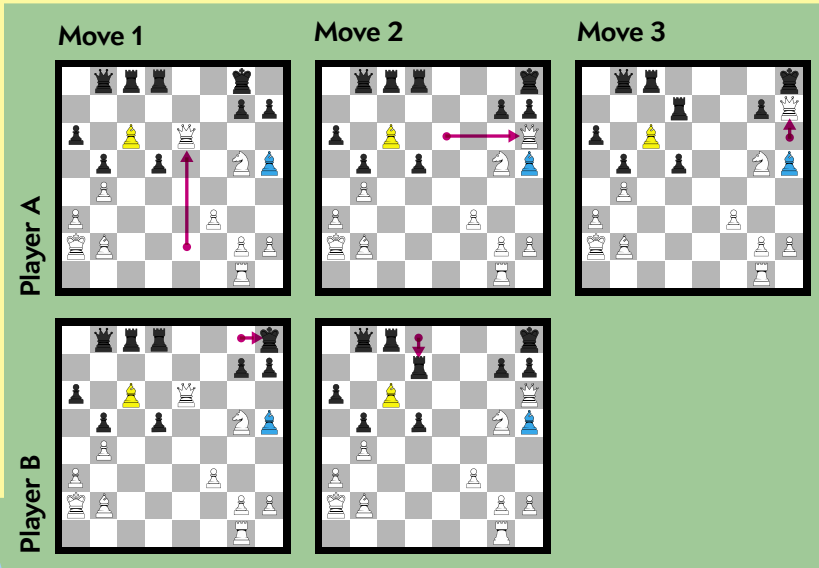
Much More Than Meets the Eye

The intellectually demanding game of chess has proved a wonderful way for psychologists to study the *Einstellung* effect—the brain's tendency to stick with solutions it already knows rather than looking for potentially superior ones. Experiments have shown that this cognitive bias literally changes how even expert chess players see the board in front of them.

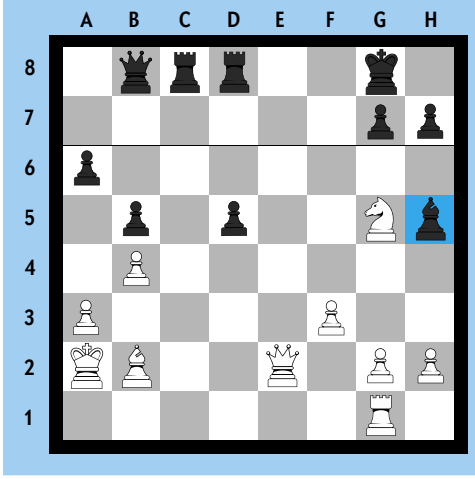


Two-Solution Problem





One-Solution Problem

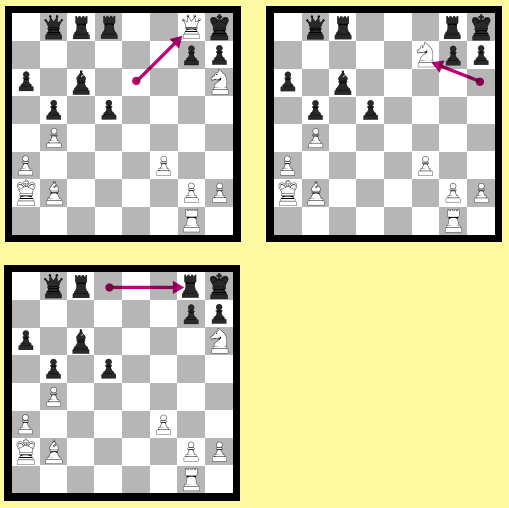


Chess Masters Fail to See the Quickest Path to Victory

In a well-known five-sequence move called smothered mate (*yellow*), player A begins by moving the queen from E2 to E6, backing player B's king into a corner. Player A then repeatedly threatens to take B's king with a knight, forcing player B to dodge. As an act of deliberate sacrifice, player A moves the queen adjacent to B's king, allowing player B to take the queen with a rook. To end the game, player A moves the knight to F7, boxing in B's king with no chance of escape. In recent experiments, psychologists presented master chess players with the two-solution board shown, which could be won using either the smothered mate or a much swifter three-step solution (*green*). The players were told to achieve checkmate as quickly as possible, but once they recognized the smothered mate as a possibility, they became seemingly incapable of noticing the more efficient strategy. When presented with a nearly identical board on which the position of one bishop had shifted (*blue*), eliminating the smothered mate as an option, the players did recognize the speedier solution, however.

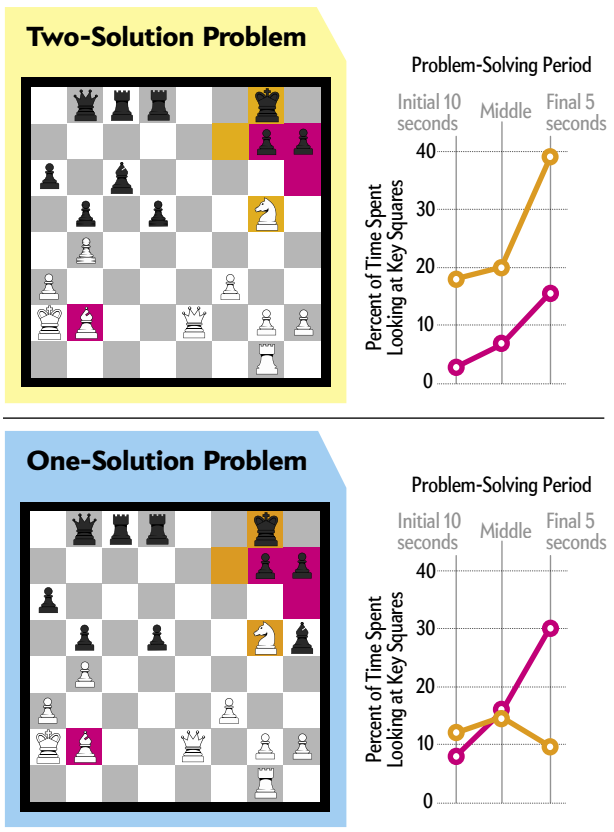
Illustration by George Retseck

© 2020 Scientific American



The Explanation: Tunnel Vision

Eye-tracking devices revealed that as soon as chess players hit on the smothered mate as a solution, they spent far more time looking at squares relevant to that familiar maneuver (*orange*) than at squares pertinent to the more efficient three-step sequence (*magenta*), despite insisting that they were searching for alternatives. Conversely, when the smothered mate was not viable, the players' gaze shifted to regions of the chessboard crucial to the swifter strategy.



that should be obvious, such as a swelling that could indicate cancer. If those secondary details are presented alone, however, radiologists see them right away.

Related studies have revealed that jurors begin to decide whether someone is innocent or guilty long before all the evidence has been presented. In turn, their initial impressions of the defendant change how they weigh subsequent evidence and even their memory of evidence they saw before. Likewise, if an interviewer finds a candidate to be physically attractive, he or she will automatically perceive that person's intelligence and personality in a more positive light, and vice versa. These biases, too, are driven by the *Einstellung* effect. It is easier to make a decision about someone if one maintains a consistent view of that person rather than sorting through contradictory evidence.

Can we learn to resist the *Einstellung* effect? Perhaps. In our chess experiments and the follow-up experiments by Sheridan and Reingold, some exceptionally skilled experts, such as grand masters, did in fact spot the less obvious optimal solution even when a slower but more familiar sequence of moves was possible. This suggests that the more expertise someone has in their field—whether chess, science or medicine—the more immune they are to cognitive bias.

But no one is completely impervious; even the grand masters failed when we made the situation tricky enough. Actively remembering that you are susceptible to the *Einstellung* effect is another way to counteract it. When considering the evidence on, say, the relative contribution of human-made and naturally occurring greenhouse gases to global temperature, remember that if you already think you know the answer, you will not judge the evidence objectively. Instead you will notice evidence that supports the opinion you already hold, evaluate it as stronger than it really is and find it more memorable than evidence that does not support your view.

We must try and learn to accept our errors if we sincerely want to improve our ideas. English naturalist Charles Darwin came up with a remarkably simple and effective technique to do just this. "I had ... during many years, followed a golden rule, namely, that whenever a published fact, a new observation or thought came across me, which was opposed by my general results, to make a memorandum of it without fail and at once," he wrote. "For I had found by experience that such facts and thoughts were far more apt to escape from memory than favourable ones." SA

Merim Bilalić is a professor of cognitive psychology at the University of Northumbria at Newcastle. His research on the *Einstellung* effect won the British Psychological Society's Award for Outstanding Doctoral Research Contributions to Psychology in 2008. His latest book is *The Neuroscience of Expertise* (Cambridge University Press, 2017).

Peter McLeod is an emeritus fellow at Queen's College at the University of Oxford. He is a trustee of the Oxford Foundation for Theoretical Neuroscience and Artificial Intelligence.

MORE TO EXPLORE

The Einstellung Effect in Anagram Problem Solving: Evidence from Eye Movements. J. J. Ellis and E. M. Reingold in *Frontiers in Psychology,* Vol. 5, Article 679; July 2014.
Baboons (*Papio papio*), but Not Humans, Break Cognitive Set in a Visuomotor Task. S. M. Pope et al. in *Animal Cognition,* Vol. 18, No. 6, pages 1339–1346; November 2015.
Magicians Fix Your Mind: How Unlikely Solutions Block Obvious Ones. C. Thomas and A. Didierjean in *Cognition,* Vol. 154, pages 169–173; September 2016.

scientificamerican.com/magazine/sa

© 2020 Scientific American

How to Think about "Implicit Bias"

Amid a controversy, it's important to remember that implicit bias is real—and it matters

By Keith Payne, Laura Niemi and John M. Doris

WHEN IS THE LAST TIME A STEREOTYPE popped into your mind? If you are like most people, the authors included, it happens all the time. That doesn't make you a racist, sexist or whatever-ist. It just means your brain is working properly, noticing patterns and making generalizations. But the same thought processes that make people smart can also make them biased. This tendency for stereotype-confirming thoughts to pass spontaneously through our minds is what psychologists call implicit bias. It sets people up to overgeneralize, sometimes leading to discrimination even when people feel they are being fair.

Studies of implicit bias have recently drawn ire from both right and left. For the right, talk of implicit bias is just another instance of progressives seeing injustice under every bush. For the left, implicit bias diverts attention from more damaging instances of explicit bigotry. Debates have become heated and have leaped from scientific journals to the popular press. Along the way, some important points have been lost. We highlight two misunderstandings that anyone who wants to understand implicit bias should know about.

First, much of the controversy centers on the most famous implicit bias test, the Implicit Association Test (IAT). A majority of people taking this test show evidence of implicit bias, suggesting that most individuals are implicitly biased even if they do not think of themselves as prejudiced. As with any measure, the test does have limitations. The stability of the test is low, meaning that if you take the same test a few weeks apart, you might score very differently. And the correlation between a person's IAT scores and discriminatory behavior is often small.

The IAT is a measure, and it doesn't follow from a particular *measure* being flawed that the *phenomenon* we are attempting to measure is not real. Drawing that conclusion is to commit the *Divining Rod Fallacy*: just because a rod doesn't find water doesn't mean there's no such thing as water. A smarter move is to ask, "What does the other evidence show?"

In fact, there is lots of other evidence. There are perceptual illusions, for example, in which white subjects perceive Black faces as angrier than white faces with the same expression. Race can bias people to see harmless objects as weapons when they are in the hands of Black men and to dislike abstract images that are paired with Black faces. And there are dozens of variants of laboratory tasks finding that most participants are faster to identify bad words paired with Black faces than white faces. None of these measures is without limitations, but they show the same pattern of reliable bias as the IAT. There is a mountain of evidence—independent of any single test—that implicit bias is real.

The second misunderstanding is about what scientists mean when they say a measure predicts behavior. One frequent complaint is that an individual's IAT score doesn't tell you whether the person will discriminate on a particular occasion. This is to commit the *Palm Reading Fallacy:* unlike palm readers, research psychologists aren't usually in the business of telling you, as an individual, what your life holds in store. Most measures in psychology, from aptitude tests to personality scales, are useful for predicting how *groups* will respond *on average,* not forecasting how particular *individuals* will behave.

The difference is crucial. Knowing that

© 2020 Scientific American

LYUBOV IVANOVA *Getty Images*

an employee scored high on conscientiousness won't tell you much about whether her work will be careful or sloppy if you inspect it right now. But if a large company hires hundreds of employees who are all conscientious, this will likely pay off with a small but consistent increase in careful work on average.

Implicit bias researchers have always warned against using the tests for predicting individual outcomes, such as how a particular manager will behave in job interviews—they've never been in the palm-reading business. What the IAT does, and does well, is predict average outcomes across larger entities like counties, cities or states. For example, metro areas with greater average implicit bias have larger racial disparities in police shootings. And counties with greater average implicit bias have larger racial disparities in infant health problems. These correlations are important: the lives of Black citizens and newborn Black babies depend on them.

Field experiments demonstrate that real-world discrimination continues and is widespread. White applicants get about 50 percent more call-backs than Black applicants with the same resumes; college professors are 26 percent more likely to respond to a student's e-mail when it is signed by Brad rather than Lamar; and physicians recommend less pain medication for Black patients than white patients with the same injury.

Today managers are unlikely to announce that white job applicants should be chosen over Black applicants, and physicians don't declare that Black people feel less pain than whites. Yet the widespread pattern of discrimination and disparities seen in field studies persists. It bears a much closer resemblance to the widespread stereotypical thoughts seen on implicit tests than to the survey studies in which most people present themselves as unbiased.

One reason people on both the right and the left are skeptical of implicit bias might be pretty simple: it isn't nice to think we aren't very nice. It would be comforting to conclude, when we don't consciously entertain impure intentions, that all of our intentions are pure. Unfortunately, we can't conclude that: many of us are more biased than we realize. And that is an important cause of injustice—whether you know it or not. SA

Keith Payne is a professor in psychology and neuroscience at the University of North Carolina at Chapel Hill. He is author of *The Broken Ladder: How Inequality Affects the Way We Think, Live, and Die* (Viking, 2017).

Laura Niemi is an assistant professor in the department of psychology at Cornell University. She studies moral judgment and the implications of differences in moral values.

John M. Doris is Peter L. Dyson Professor of Ethics in Organizations and Life at the Charles H. Dyson School of Applied Economics and Management and a professor at the Sage School of Philosophy at Cornell University.

MORE TO EXPLORE

Physicians and Implicit Bias: How Doctors May Unwittingly Perpetuate Health Care Disparities. Elizabeth N. Chapman, Anna Kaatz and Molly Carnes in *Journal of General Internal Medicine,* Vol. 28, No. 11, pages 1504–1510; November 2013.

Condoning Stereotyping? How Awareness of Stereotyping Prevalence Impacts Expression of Stereotypes. Michelle M. Duguid and Melissa C. Thomas-Hunt in *Journal of Applied Psychology,* Vol. 100, No. 2, pages 343–359; March 2015.

scientificamerican.com/magazine/sa

© 2020 Scientific American

Why We Trust Lies

The most effective misinformation starts with seeds of truth

By Cailin O'Connor and James Owen Weatherall

IN THE MID-1800S A CATERPILLAR THE SIZE OF A HUMAN FINGER began spreading across the northeastern U.S. This appearance of the tomato hornworm was followed by terrifying reports of fatal poisonings and aggressive behavior toward people. In July 1869 newspapers across the region posted warnings about the insect, reporting that a girl in Red Creek, N.Y., had been "thrown into spasms, which ended in death" after a run-in with the creature. That fall the *Syracuse Standard* printed an account from one Dr. Fuller, who had collected a particularly enormous specimen. The physician warned that the caterpillar was "as poisonous as a rattlesnake" and said he knew of three deaths linked to its venom.

Although the hornworm is a voracious eater that can strip a tomato plant in a matter of days, it is, in fact, harmless to humans. Entomologists had known the insect to be innocuous for decades when Fuller published his dramatic account, and his claims were widely mocked by experts. So why did the rumors persist even though the truth was readily available? People are social learners. We develop most of our beliefs from the testimony of trusted others such as our teachers, parents and friends. This social transmission of knowledge is at the heart of culture and science. But as the tomato hornworm story shows us, our ability has a gaping vulnerability: sometimes the ideas we spread are wrong.

Over the past five years the ways in which the social transmission of knowledge can fail us have come into sharp focus. Misinformation shared on social media Web sites has fueled an epidemic of false belief, with widespread misconceptions concerning topics ranging from the COVID-19 pandemic, to voter fraud, whether the Sandy Hook school shooting was staged, and whether vaccines are safe. The same basic mechanisms that spread fear about the tomato hornworm have now intensified—and, in some cases, led to—a profound public mistrust of basic societal institutions. One consequence is the largest measles outbreak in a generation.

"Misinformation" may seem like a misnomer here. After all, many of today's most damaging false beliefs are initially driven by acts of propaganda and disinformation, which are deliberately deceptive and intended to cause harm. But part of what makes disinformation so effective in an age of social media is the fact that people who are exposed to it share it widely among friends and peers who trust them, with no intention of misleading anyone. Social media transforms disinformation into misinformation.

Many communication theorists and social scientists have tried to understand how false beliefs persist by modeling the spread of ideas as a contagion. Employing mathematical models involves simulating a simplified representation of human social interactions using a computer algorithm and then studying these simulations to learn something about the real world. In a contagion model, ideas are like viruses that go from mind to mind. You start with a network, which consists of nodes, representing individuals, and edges, which represent social connections. You seed an idea in one "mind" and see how it spreads under various assumptions about when transmission will occur.

Contagion models are extremely simple but have been used to explain surprising patterns of behavior, such as the epidemic of suicide that reportedly swept through Europe after publication of Goethe's *The Sorrows of Young Werther* in 1774 or when dozens of U.S. textile workers in 1962 reported suffering from nausea and numbness after being bitten by an imaginary insect. They can also explain how some false beliefs propagate on the Internet. Before the last U.S. presidential election, an image of a young Donald Trump appeared on Facebook. It included a quote, attributed to a 1998 interview in *People* magazine, saying that if Trump ever ran for president, it would be as a Republican because the party is made up of "the dumbest group of voters." Although it is unclear who "patient zero" was, we know that this meme passed rapidly from profile to profile.

The meme's veracity was quickly evaluated and debunked. The fact-checking Web site Snopes reported that the quote was fabricated as early as October 2015. But as with the tomato hornworm, these efforts to disseminate truth did not change how the rumors spread. One copy of the meme alone was shared more than half a million times. As new individuals shared it over the next several years, their false beliefs infected friends who observed the meme, and they, in turn, passed the false belief on to new areas of the network.

This is why many widely shared memes seem to be immune to fact-checking and debunking. Each person who shared the Trump meme simply trusted the friend who had shared it rather than checking for themselves. Putting the facts out there does not help if no one bothers to look them up. It might seem like the problem here is laziness or gullibility—and thus that the solution is merely more education or better critical thinking skills. But that is not entirely right. Sometimes false beliefs persist and spread even in communities where everyone works very hard to learn the truth by gathering and sharing evidence. In these cases, the problem is not unthinking trust. It goes far deeper than that.

IN BRIEF

Social media has facilitated the proliferation of false belief at an unprecedented scale. **By modeling** the ways misinformation spreads via networks of people, researchers learn how social trust and conformity affect how communities reach consensus. **Adding propagandists** to the models shows how easily belief can be manipulated, even when scientists collect ample evidence.

© 2020 Scientific American

TRUST THE EVIDENCE

THE FACEBOOK PAGE "Stop Mandatory Vaccination" has more than 140,000 followers. Its moderators regularly post material that is framed to serve as evidence for this community that vaccines are harmful or ineffective, including news stories, scientific papers and interviews with prominent vaccine skeptics. On other Facebook group pages, thousands of concerned parents ask and answer questions about vaccine safety, often sharing scientific papers and legal advice supporting antivaccination efforts. Participants in these online communities care very much about whether vaccines are harmful and actively try to learn the truth. Yet they come to dangerously wrong conclusions. How does this happen?

The contagion model is inadequate for answering this question. Instead we need a model that can capture cases where people form beliefs on the basis of evidence that they gather and share. It must also capture why these individuals are motivated to seek the truth in the first place. When it comes to health topics, there might be serious costs to acting on false beliefs. If vaccines are safe and effective (which they are) and parents do not vaccinate, they put their kids and immunosuppressed people at unnecessary risk. If vaccines are not safe, as the participants in these Facebook groups have concluded, then the risks go the other way. This means that figuring out what is true, and acting accordingly, matters deeply.

To better understand this behavior in our research, we drew

Illustration by Lisk Feng

© 2020 Scientific American

How Network Science Maps the Spread of Misinformation

We use network science to better understand how social connections influence the beliefs and behaviors of individuals in a social network—and especially how false beliefs can spread from person to person. Here we look at two kinds of network models that capture different ways in which ideas or beliefs spread. Each node in these models represents an individual. Each edge, or connection between the nodes, represents a social tie.

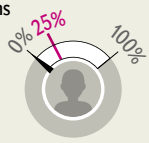
THE CONTAGION MODEL

Contagion models treat ideas or beliefs like viruses that spread between individuals in a social network. There are different ways that this "infection" can work. In some models, everyone will be infected by an infected neighbor. In others, ideas spread whenever some percentage of an individual's neighbors become infected. Here we illustrate these "complex contagions" with examples where individuals take on a new belief if at least 25 percent of their neighbors hold it. In these models, the structure of the network affects how ideas spread.

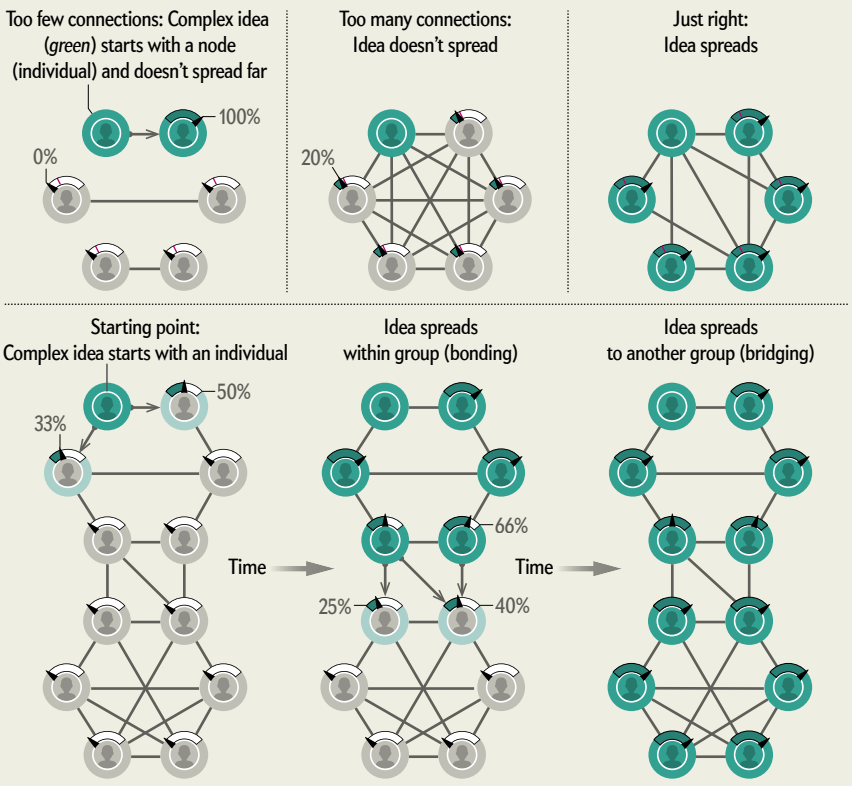
How to Read the Contagion Plots

Each circular node is a person who is influenced by ideas presented by others. Each line represents a connection between individuals.

The gauge at the top of the nodes indicates the percent of that person's connections who hold a particular belief. In the scenarios below, the threshold for an individual to take on the belief of their neighbors is 25% (at least 1 out of 4).



BONDING AND BRIDGING: In less connected groups, ideas cannot reach all members. Sometimes too many connections can also stop the spread of an idea. Some networks have tight-knit cliques, where even if an idea spreads within one clique, it can be difficult for it to spread to other cliques.



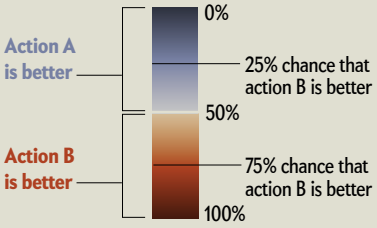
NETWORK EPISTEMOLOGY FRAMEWORK

Network epistemology models represent situations in which people form beliefs by gathering and sharing evidence. This kind of model applies to many cases in science. Beliefs do not simply spread from individual to individual. Instead each individual has some degree of certainty about an idea. This prompts the person to gather evidence in support of it, and that evidence changes beliefs. Each individual shares the evidence with network neighbors, which also influences beliefs.

How to Read the Network Epistemology Framework Plots

Each circular or square node is a person who is influenced by evidence presented by others. Each has a belief about whether **action A** (*blue*) or **action B** (*orange*) is better. The person's belief can strengthen, weaken and/or flip over time, as shown here by changing colors.

The strength of the color represents the individual's level of certainty in a particular action. For example, an assignment of 75% means that the individual thinks that there is a 75% chance that **action B** is better than **action A**. If the value is greater than 50%, then the individual performs **action B**. Then, we use Bayes's rule—which probability theory tells us is the rational way to change beliefs in light of evidence—to update the individual's credence in light of this result and then update all connections in his or her network.



Square nodes are individuals who test actions and update their beliefs accordingly (evidence seekers)

Circular nodes represent individuals who observe results from others but do not test the actions directly (observers)

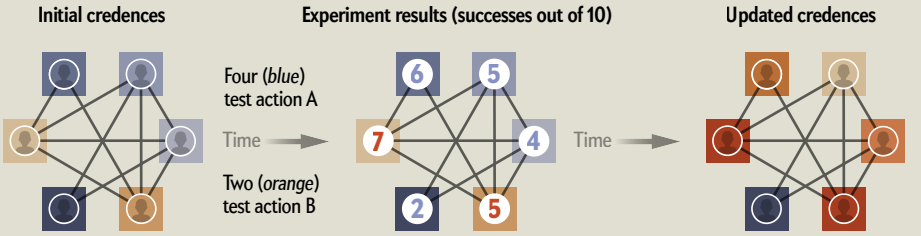
Stars represent individuals who do not hold beliefs of their own but instead focus on introducing selective results into the system (propagandists)

SOURCES: THE WISDOM AND/OR MADNESS OF CROWDS, BY NICKY CASE *https://ncase.me/crowds*; *THE MISINFORMATION AGE: HOW FALSE BELIEFS SPREAD*, BY CAILIN O'CONNOR AND JAMES OWEN WEATHERALL. YALE UNIVERSITY PRESS, 2019

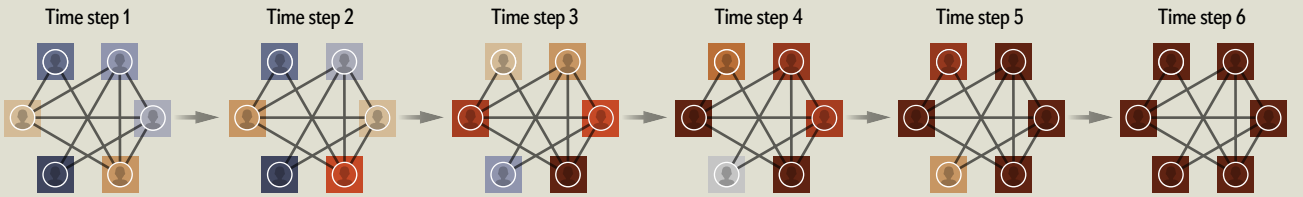
Graphic by Jen Christiansen

© 2020 Scientific American

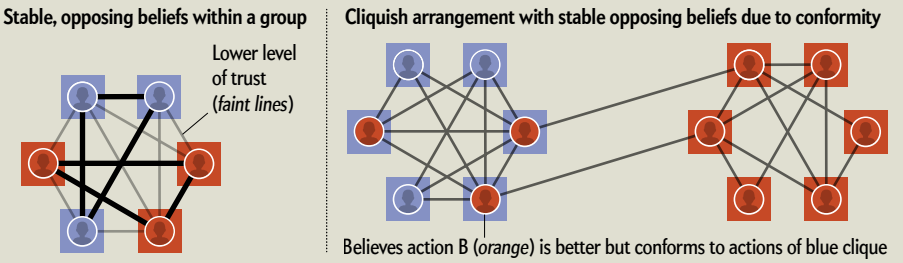
UPDATING AND EXPERIMENTING: Individuals in these models start with some random level of certainty, or credence, about whether action A or B is better. They then take the action they prefer—that is, "experiment." Their outcomes provide evidence about the success of these actions, which they share with neighbors. All individuals update their credences based on what they observe.



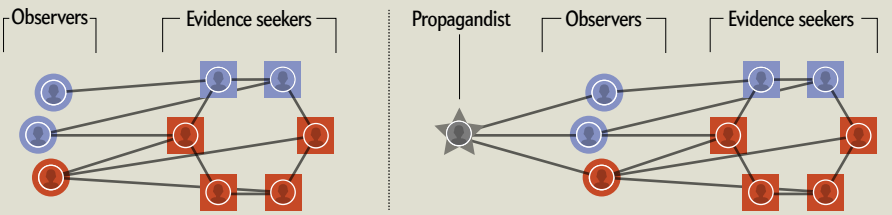
CONVERGENCE ON TRUE BELIEFS: Over time, the social connections in these models mean that groups of people come to a consensus about whether A or B is better. As they gather and share evidence, they usually learn that the better action is, indeed, better. Someone trying the worse action, for instance, will see how much better their neighbor is doing and switch. Sometimes, though, strings of misleading evidence will convince the entire group that the worse action is better.



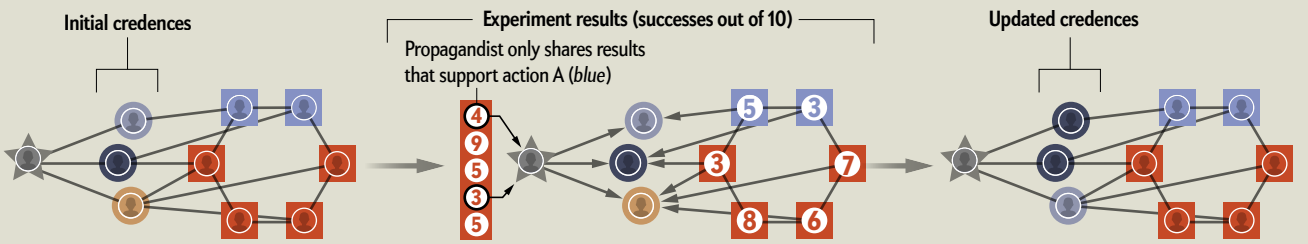
POLARIZATION: If we add social trust or conformity to these models, they may no longer reach consensus. If each individual trusts the evidence that comes from those who share their beliefs, polarized camps that only listen to those in their group form. If each individual seeks to conform their actions with group members, good ideas fail to spread between cliques.


Stable, opposing beliefs within a group
Lower level of trust (faint lines)
Cliquish arrangement with stable opposing beliefs due to conformity
Believes action B (orange) is better but conforms to actions of blue clique

EVIDENCE SEEKERS, OBSERVERS AND PROPAGANDISTS: In some cases, propagandists try to mislead a group of people about scientific results. We can use these models to represent a set of evidence seekers who gather evidence, a group of observers who update beliefs based on this evidence, and a propagandist who misleads the observers.



BELIEF UPDATING WHEN SELECTIVE RESULTS ARE IN PLAY: Industrial propagandists shape public belief by selectively sharing only those results that happen to spuriously support the worse action. This can mislead the public, even in cases when groups of evidence seekers converge to a consensus about the true belief. This strategy for public disinformation takes advantage of the inherent randomness of scientific results to mislead.



© 2020 Scientific American

on what is called the network epistemology framework. It was first developed by economists 20 years ago to study the social spread of beliefs in a community. Models of this kind have two parts: a problem and a network of individuals (or "agents"). The problem involves picking one of two choices: These could be "vaccinate" and "don't vaccinate" your children. In the model, the agents have beliefs about which choice is better. Some believe vaccination is safe and effective, and others believe it causes autism. Agent beliefs shape their behavior—those who think vaccination is safe choose to perform vaccinations. Their behavior, in turn, shapes their beliefs. When agents vaccinate and see that nothing bad happens, they become more convinced vaccination is indeed safe.

The second part of the model is a network that represents social connections. Agents can learn not only from their own experiences of vaccinating but also from the experiences of their neighbors. Thus, an individual's community is highly important in determining what beliefs they ultimately develop.

The network epistemology framework captures some essential features missing from contagion models: individuals intentionally gather data, share data and then experience consequences for bad beliefs. The findings teach us some important lessons about the social spread of knowledge. The first thing we learn is that working together is better than working alone, because an individual facing a problem like this is likely to prematurely settle on the worse theory. For instance, he or she might observe one child who turns out to have autism after vaccination and conclude that vaccines are not safe. In a community, there tends to be some diversity in what people believe. Some test one action; some test the other. This diversity means that usually enough evidence is gathered to form good beliefs.

But even this group benefit does not *guarantee* that agents learn the truth. Real scientific evidence is probabilistic, of course. For example, some nonsmokers get lung cancer, and some smokers do not get lung cancer. This means that some studies of smokers will find no connection to cancer. Relatedly, although there is no actual statistical link between vaccines and autism, some vaccinated children will be autistic. Thus, some parents observe their children developing symptoms of autism after receiving vaccinations. Strings of misleading evidence of this kind can be enough to steer an entire community wrong.

In the most basic version of this model, social influence means that communities end up at consensus. They decide either that vaccinating is safe or that it is dangerous. But this does not fit what we see in the real world. In actual communities, we see polarization—entrenched disagreement about whether or not to vaccinate. We argue that the basic model is missing two crucial ingredients: social trust and conformism.

PROTESTERS use the language of "choice" to spread misinformation about vaccine safety.

Social trust matters to belief when individuals treat some sources of evidence as more reliable than others. This is what we see when anti-vaxxers trust evidence shared by others in their community more than evidence produced by the Centers for Disease Control and Prevention or other medical research groups. This mistrust can stem from all sorts of things, including previous negative experiences with doctors or concerns that health care or governmental institutions do not care about their best interests. In some cases, this distrust may be justified, given that there is a long history of medical researchers and clinicians ignoring legitimate issues from patients, particularly women.

Yet the net result is that anti-vaxxers do not learn from the very people who are collecting the best evidence on the subject. In versions of the model where individuals do not trust evidence from those who hold very different beliefs, we find communities polarize, and those with poor beliefs fail to learn better ones.

Conformism, meanwhile, is a preference to act in the same way as others in one's community. The urge to conform is a profound part of the human psyche and one that can lead us to take actions we know to be harmful. When we add conformism to the model, what we see is the emergence of cliques of agents who hold false beliefs. The reason is that agents connected to the outside world do not pass along information that conflicts with their group's beliefs, meaning that many members of the group never learn the truth.

Conformity can help explain why vaccine skeptics tend to cluster in certain communities. Some private and charter schools in southern California have vaccination rates in the low double digits. And rates are startlingly low among Somali immigrants in Minneapolis and Orthodox Jews in Brooklyn—two communities that have recently suffered from measles outbreaks.

Interventions into vaccine skepticism need to be sensitive to both social trust and conformity. Simply sharing new evidence with skeptics will likely not help because of trust issues. And convincing trusted community members to speak out for vaccination might be difficult because of conformism. The best approach is to find individuals who share enough in common with members of the relevant communities to establish trust. A rabbi, for instance, might be an effective vaccine ambassador in Brooklyn, whereas in southern California, you might need to get Gwyneth Paltrow involved.

Social trust and conformity can help explain why polarized beliefs can emerge in social networks. But at least in some cases, including the Somali community in Minnesota and Orthodox Jewish communities in New York, they are only part of the story. Both groups were the targets of sophisticated misinformation campaigns designed by anti-vaxxers.

RICH PEDRONCELLI *AP Photo*

© 2020 Scientific American

INFLUENCE OPERATIONS

HOW WE VOTE, what we buy and who we acclaim all depend on what we believe about the world. As a result, there are many wealthy, powerful groups and individuals who are interested in shaping public beliefs—including those about scientific matters of fact. There is a naive idea that when industry attempts to influence scientific belief, they do it by buying off corrupt scientists. Perhaps this happens sometimes. But a careful study of historical cases shows there are much more subtle—and arguably more effective—strategies that industry, nation states and other groups utilize. The first step in protecting ourselves from this kind of manipulation is to understand how these campaigns work.

A classic example comes from the tobacco industry, which developed new techniques in the 1950s to fight the growing consensus that smoking kills. During the 1950s and 1960s the Tobacco Institute published a bimonthly newsletter called "Tobacco and Health" that reported only scientific research suggesting tobacco was not harmful or research that emphasized uncertainty regarding the health effects of tobacco.

The pamphlets employ what we have called selective sharing. This approach involves taking real, independent scientific research and curating it, by presenting only the evidence that favors a preferred position. Using variants on the models described earlier, we have argued that selective sharing can be shockingly effective at shaping what an audience of nonscientists comes to believe about scientific matters of fact. In other words, motivated actors can use seeds of truth to create an impression of uncertainty or even convince people of false claims.

Selective sharing has been a key part of the anti-vaxxer playbook. Before the recent measles outbreak in New York, an organization calling itself Parents Educating and Advocating for Children's Health (PEACH) produced and distributed a 40-page pamphlet entitled "The Vaccine Safety Handbook." The information shared—when accurate—was highly selective, focusing on a handful of scientific studies suggesting risks associated with vaccines, with minimal consideration of the many studies that find vaccines to be safe.

The PEACH handbook was especially effective because it combined selective sharing with rhetorical strategies. It built trust with Orthodox Jews by projecting membership in their community (though published pseudonymously, at least some authors *were* members) and emphasizing concerns likely to resonate with them. It cherry-picked facts about vaccines intended to repulse its particular audience; for instance, it noted that some vaccines contain gelatin derived from pigs. Wittingly or not, the pamphlet was designed in a way that exploited social trust and conformism—the very mechanisms crucial to the creation of human knowledge.

Worse, propagandists are constantly developing ever more sophisticated methods for manipulating public belief. Over the past several years we have seen purveyors of disinformation roll out new ways of creating the impression—especially through social media conduits such as Twitter bots and paid trolls and, most recently, by hacking or copying your friends' accounts—that certain false beliefs are widely held, including by your friends and others with whom you identify. Even the PEACH creators may have encountered this kind of synthetic discourse about vaccines. According to a 2018 article in the *American Journal of Public Health*, such disinformation was distributed by accounts linked to Russian influence operations seeking to amplify American discord and weaponize a public health issue. This strategy works to change minds not through rational arguments or evidence but simply by manipulating the social spread of knowledge and belief.

The sophistication of misinformation efforts (and the highly targeted disinformation campaigns that amplify them) raises a troubling problem for democracy. Returning to the measles example, children in many states can be exempted from mandatory vaccinations on the grounds of "personal belief." This became a flash point in California in 2015 following a measles outbreak traced to unvaccinated children visiting Disneyland. Then governor Jerry Brown signed a new law, SB277, removing the exemption.

Immediately vaccine skeptics filed paperwork to put a referendum on the next state ballot to overturn the law. Had they succeeded in getting 365,880 signatures (they made it to only 233,758), the question of whether parents should be able to opt out of mandatory vaccination on the grounds of personal belief would have gone to a direct vote—the results of which would have been susceptible to precisely the kinds of disinformation campaigns that have caused vaccination rates in many communities to plummet.

Luckily, the effort failed. But the fact that hundreds of thousands of Californians supported a direct vote about a question with serious bearing on public health, where the facts are clear but widely misconstrued by certain activist groups, should give serious pause. There is a reason that we care about having policies that best reflect available evidence and are responsive to reliable new information. How do we protect public well-being when so many citizens are misled about matters of fact? Just as individuals acting on misinformation are unlikely to bring about the outcomes they desire, societies that adopt policies based on false belief are unlikely to get the results they want and expect.

The way to decide a question of scientific fact—are vaccines safe and effective?—is not to ask a community of nonexperts to vote on it, especially when they are subject to misinformation campaigns. What we need is a system that not only respects the processes and institutions of sound science as the best way we have of learning the truth about the world but also respects core democratic values that would preclude a single group, such as scientists, dictating policy.

We do not have a proposal for a system of government that can perfectly balance these competing concerns. But we think the key is to better separate two essentially different issues: What are the facts, and what should we do in light of them? Democratic ideals dictate that both require public oversight, transparency and accountability. But it is only the second—how we should make decisions given the facts—that should be up for a vote. SA

Cailin O'Connor is an associate professor of logic and philosophy of science and **James Owen Weatherall** is a professor of logic and philosophy of science at the University of California, Irvine.

They are co-authors of *The Misinformation Age: How False Beliefs Spread* (Yale University Press, 2019). Both are members of the Institute for Mathematical Behavioral Sciences.

MORE TO EXPLORE

Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate. David A. Broniatowski et al. in *American Journal of Public Health*, Vol. 108, No. 10, pages 1378–1384; October 2018.

The Wisdom and/or Madness of Crowds. Nicky Case. Interactive game for network contagion: https://ncase.me/crowds

scientificamerican.com/magazine/sa

© 2020 Scientific American

TOUGH CALLS

How we make decisions in the face of incomplete knowledge and uncertainty

By Baruch Fischhoff

Illustration by Wesley Allsbrook

© 2020 Scientific American

© 2020 Scientific American

PSYCHOLOGISTS STUDY HOW HUMANS MAKE DECISIONS by giving people "toy" problems. In one study, for example, my colleagues and I described to subjects a hypothetical disease with two strains. Then we asked, "Which would you rather have? A vaccine that completely protects you against one strain or a vaccine that gives you 50 percent protection against both strains?" Most people chose the first vaccine. We inferred that they were swayed by the phrase about complete protection, even though both shots gave the same overall chance of getting sick.

IN BRIEF

When people assess novel risks, they rely on mental models derived from previous experience, which may not be applicable. **Asking people** how they form such assessments can reveal misleading preconceptions. **Experts can** also test messages about risk to ensure the public understands them clearly.

But we live in a world with real problems, not just toy ones—situations that sometimes require people to make life-and-death decisions in the face of incomplete or uncertain knowledge. Years ago, after I had begun to investigate decision-making with my colleagues Paul Slovic and the late Sarah Lichtenstein, both at the firm Decision Research in Eugene, Ore., we started getting calls about non-toy issues—calls from leaders in industries that produced nuclear power or genetically modified organisms (GMOs). The gist was: "We've got a wonderful technology, but people don't like it. Even worse, they don't like us. Some even think that we're evil. You're psychologists. Do something."

We did, although it probably wasn't what these company officials wanted. Instead of trying to change people's minds, we set about learning how they *really* thought about these technologies. To that end, we asked them questions designed to reveal how they assessed risks. The answers helped us understand why people form beliefs about divisive issues such as nuclear energy—and today, climate change—when they do not have all the facts.

INTIMATIONS OF MORTALITY

TO START OFF, we wanted to figure out how well the general public understands the risks they face in everyday life. We asked groups of laypeople to estimate the annual death toll from causes such as drowning, emphysema and homicide and then compared their estimates with scientific ones. Based on previous research, we expected that people would make generally accurate predictions but that they would overestimate deaths from causes that get splashy or frequent headlines—murders, tornadoes—and underestimate deaths from "quiet killers," such as stroke and asthma, that do not make big news as often.

Overall, our predictions fared well. People overestimated highly reported causes of death and underestimated ones that received less attention. Images of terror attacks, for example, might explain why people who watch more television news worry more about terrorism than individuals who rarely watch. But one puzzling result emerged when we probed these beliefs. People who were strongly opposed to nuclear power believed that it had a very low annual death toll. Why, then, would they be against it? The apparent paradox made us wonder if by asking them to predict average annual death tolls, we had defined risk too narrowly. So, in a new set of questions we asked what risk really meant to people. When we did, we found that those opposed to nuclear power thought the technology had a greater potential to cause widespread catastrophes. That pattern held true for other technologies as well.

To find out whether knowing more about a tech-

© 2020 Scientific American

nology changed this pattern, we asked technical experts the same questions. The experts generally agreed with laypeople about nuclear power's death toll for a typical year: low. But when they defined risk themselves, on a broader time frame, they saw less potential for problems. The general public, unlike the experts, emphasized what could happen in a very bad year. The public and the experts were talking past each other and focusing on different parts of reality.

UNDERSTANDING RISK

DID EXPERTS always have an accurate understanding of the probabilities for disaster? Experts analyze risks by breaking complex problems into more knowable parts. With nuclear power, the parts might include the performance of valves, control panels, evacuation schemes and cybersecurity defenses. With GMO crops, the parts might include effects on human health, soil chemistry and insect species.

The quality and accuracy of a risk analysis depend on the strength of the science used to assess each part. Science is fairly strong for nuclear power and GMOs. For new technologies such as self-driving vehicles, it is a different story. The components of risk could be the probability of the vehicle laser-light sensors "seeing" a pedestrian, the likelihood of a pedestrian acting predictably, and the chances of a driver taking control at the exact moment when a pedestrian is unseen or unpredictable. The physics of pulsed laser-light sensors is well understood, but how they perform in snow and gloom is not. Research on how pedestrians interact with autonomous vehicles barely exists. And studies of drivers predict that they cannot stay vigilant enough to handle infrequent emergencies.

When scientific understanding is incomplete, risk analysis shifts from reliance on established facts to expert judgment. Studies of those judgments find that they are often quite good—but only when experts get good feedback. For example, meteorologists routinely compare their probability-of-precipitation forecasts with the rain gauge at their station. Given that clear, prompt feedback, when forecasters say that there is a 70 percent chance of rain, it rains about 70 percent of the time. With new technologies such as the self-driving car or gene editing, however, feedback will be a long time coming. Until it does, we will be unsure—and the experts themselves will not know—how accurate their risk estimates really are.

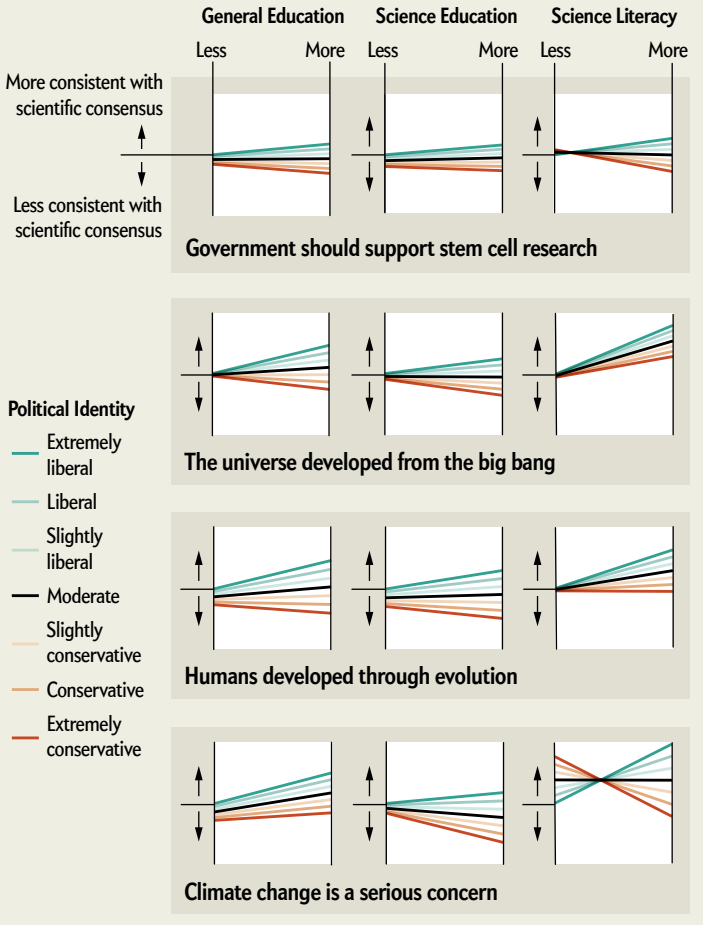
THE SCIENCE OF CLIMATE SCIENCE

EXPERT JUDGMENT, which is dependent on good feedback, comes into play when one is predicting the costs and benefits of attempts to slow climate change or to adapt to it. Climate analyses combine the judgments of experts from many research areas, including obvious ones, such as atmospheric chemistry and oceanography, and less obvious ones, such as botany, archaeology and glaciology. In complex climate analyses, these expert judgments reflect great knowedge driven by evidence-based feedback. But some aspects still remain uncertain.

My first encounter with these analyses was in 1979, as part of a project planning the next 20 years of climate research. Sponsored by the Department of Energy, the project had five working groups. One dealt with the oceans and polar regions, a second with the managed biosphere, a third with the less managed biosphere, and a fourth with economics and geopolitics. The fifth group, which I joined, dealt with social and institutional responses to the threat.

When the Public Disagrees about Science

On politically controversial scientific issues, polarization is greater among better-informed people. Investigators saw this effect in two national surveys in the U.S. The surveys, conducted in 2006 and 2010, combined to cover just more than 6,500 people. Participants were asked what they believed on several hot topics and whether they agreed with scientific consensus. As education and science literacy increased among liberals and conservatives, so did their divergence. This may be because more well-versed people are better attuned to the position of their political group and more confident in defending it.



SOURCE: "INDIVIDUALS WITH GREATER SCIENCE LITERACY AND EDUCATION HAVE MORE POLARIZED BELIEFS ON CONTROVERSIAL SCIENCE TOPICS," BY CAITLIN DRUMMOND AND BARUCH FISCHHOFF, IN *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES USA*, VOL. 114, NO. 36; SEPTEMBER 5, 2017

Graphic by Jen Christiansen

© 2020 Scientific American

Even then, more than 40 years ago, the evidence was strong enough to reveal the enormous gamble being taken with our planet. Our overall report, summarizing all five groups, concluded that "the probable outcome is beyond human experience."

THINKING OF THE UNTHINKABLE

HOW, THEN, CAN RESEARCHERS in this area fulfill their duty to inform people about accurate ways to think about events and choices that are beyond their experience? Scientists can, in fact, accomplish this if they follow two basic lessons from studies of decision-making.

LESSON 1: The facts of climate science will not speak for themselves. The science needs to be translated into terms that are relevant to people's decisions about their lives, their communities and their society. While most scientists are experienced communicators in a classroom, out in the world they may not get feedback on how clear or relevant their messages are.

Risky Business

The way people view the risks of technologies and activities depends on factors such as familiarity, whether exposure is voluntary or involuntary, and the likelihood of fatalities. Novelty, involuntary exposure and lethal potential lead people to rate things as riskier, assessments that sometimes differ from scientific counts and estimates. The results come from surveys given to laypeople and first published in 1978; they have been repeated often with similar findings.

Addressing this feedback problem is straightforward: test messages before sending them. One can learn a lot simply by asking people to read and paraphrase a message. When communication researchers have asked for such rephrasing about weather forecasts, for example, they have found that some are confused by the statement that there is a "70 percent chance of rain." The problem is with the words, not the number. Does the forecast mean it will rain 70 percent of the time? Over 70 percent of the area? Or there is a 70 percent chance of at least 0.01 inch of rain at the weather station? The last interpretation is the correct answer.

Many studies have found that numbers, such as 70 percent, generally communicate much better than "verbal quantifiers," such as "likely," "some" or "often." One classic case from the 1950s involves a U.S. National Intelligence Estimate that said that "an attack on Yugoslavia in 1951 should be considered a serious possibility." When asked what probability they had in mind, the analysts who signed the document gave a wide range of numbers, from 20 to 80 percent. (The Soviets did not invade.)

Sometimes people want to know more than the probability of rain or war when they make decisions. They want to understand the processes that lead to those probabilities: how things work. Studies have found that some critical aspects of climate change research are not intuitive for many people, such as how scientists can bicker yet still agree about the threat of climate change or how carbon dioxide is different from other pollutants. (It stays in the atmosphere longer.) People may reject the research results unless scientists tell them more about how they were derived.

LESSON 2: People who agree on the facts can still disagree on what to do about them. A solution that seems sound to some can seem too costly or unfair to others.

For example, people who like plans for carbon capture and sequestration, because it keeps carbon dioxide out of the air, might oppose using it on coal-fired power plants. They fear an indirect consequence: cleaner coal may make mountaintop-removal mining more acceptable. Those who know what cap-and-trade schemes are meant to do—create incentives for reducing emissions—might still believe that they will benefit banks more than the environment.

These examples show why two-way communication is so important in these situations. We need to learn what is on others' minds and make them feel like partners in decision-making. Sometimes that communication will reveal misunderstandings that research can reduce. Or it may reveal solutions that make more people happy. One example is British Columbia's revenue-neutral carbon tax, whose revenues make other taxes lower; it has also produced broad enough political support to weather several changes of government since 2008. Sometimes, of course, better two-way communication will reveal fundamental disagreements,

SOURCE: *RISK: A VERY SHORT INTRODUCTION*, BY BARUCH FISCHHOFF AND JOHN KADVANY. OXFORD UNIVERSITY PRESS, 2011; REDRAFTED FROM "HOW SAFE IS SAFE ENOUGH? A PSYCHOMETRIC STUDY OF ATTITUDES TOWARDS TECHNOLOGICAL RISKS AND BENEFITS," BY BARUCH FISCHHOFF ET AL., IN *POLICY SCIENCES*, VOL. 9, NO. 2; APRIL 1978

© 2020 Scientific American

and in those cases action becomes a matter for the courts, streets and ballot boxes.

MORE THAN SCIENCE

THESE LESSONS about how facts are communicated and interpreted are important because climate-related decisions are not always based on what research says or shows. For some individuals, scientific evidence or economic impacts are less important than what certain decisions reveal about their beliefs. These people ask how their choice will affect the way others think about them, as well as how they think about themselves.

For instance, there are people who forgo energy conservation measures but not because they are against conservation. They just do not want to be perceived as eco-freaks. Others who conserve do it more as a symbolic gesture and not based on a belief that it makes a real difference. Using surveys, researchers at Yale Climate Connections have identified what they call Six Americas in terms of attitudes, ranging from alarmed to dismissive. People at those two extremes are the ones who are most likely to adopt measures to conserve energy. The alarmed group's motives are what you might expect. Those in the dismissive group, though, may see no threat from climate change but also have noted they can save money by reducing their energy consumption.

Knowing the science does not necessarily mean agreeing with the science. The Yale study is one of several that found greater polarization among different political groups as people in the groups gained knowledge of some science-related issues. In ongoing research, Caitlin Drummond, currently a postdoctoral fellow at the University of Michigan's Erb Institute, and I have uncovered a few hints that might account for this phenomenon. One possible explanation is that more knowledgeable people are more likely to know the position of their affiliated political group on an issue and align themselves with it. A second possibility is that they feel more confident about arguing the issues. A third, related explanation is that they are more likely to see, and seize, the chance to express themselves than those who do not know as much.

WHEN DECISIONS MATTER MOST

ALTHOUGH DECISION SCIENCE researchers still have much to learn, their overall message about ways to deal with uncertain, high-stakes situations is optimistic. When scientists communicate poorly, it often indicates that they have fallen prey to a natural human tendency to exaggerate how well others understand them. When laypeople make mistakes, it often reflects their reliance on mental models that have served them well in other situations but that are not accurate in current circumstances. When people disagree about what decisions to make, it is often because they have different goals rather than different facts.

In each case, the research points to ways to help people better understand one another and themselves. Communication studies can help scientists create clearer messages. And decision science can help the public to refine their mental models to interpret new phenomena. By reducing miscommunication and focusing on legitimate disagreements, decision researchers can help society have fewer conflicts and make dealing with the ones that remain easier for us all. SA

YOUNG ACTIVISTS gathered in New York City in May 2019 to demand immediate action on climate change.

MICHAEL NIGRO *Getty Images*

Psychologist **Baruch Fischhoff** is Howard Heinz University Professor in the department of engineering and public policy and the Institute for Politics and Strategy at Carnegie Mellon University. He is a member of the National Academy of Sciences and National Academy of Medicine and past president of the Society for Risk Analysis.

MORE TO EXPLORE

Risk: A Very Short Introduction. Baruch Fischhoff and John Kadvany. Oxford University Press, 2011.
The Science of Science Communication. Special issue. *Proceedings of the National Academy of Sciences USA,* Vol. 111, Supplement 3; August 20, 2013. www.pnas.org/content/110/Supplement_3
The Science of Science Communication II. Special issue. *Proceedings of the National Academy of Sciences USA,* Vol. 111, Supplement 4; September 16, 2014. www.pnas.org/content/111/Supplement_4
The Science of Science Communication III: Inspiring Novel Collaborations and Building Capacity: Proceedings of a Colloquium. National Academy of Sciences. National Academies Press, 2018.

scientificamerican.com/magazine/sa

© 2020 Scientific American

Confronting Unknowns

How to interpret uncertainty in common forms of data visualization

By Jessica Hullman

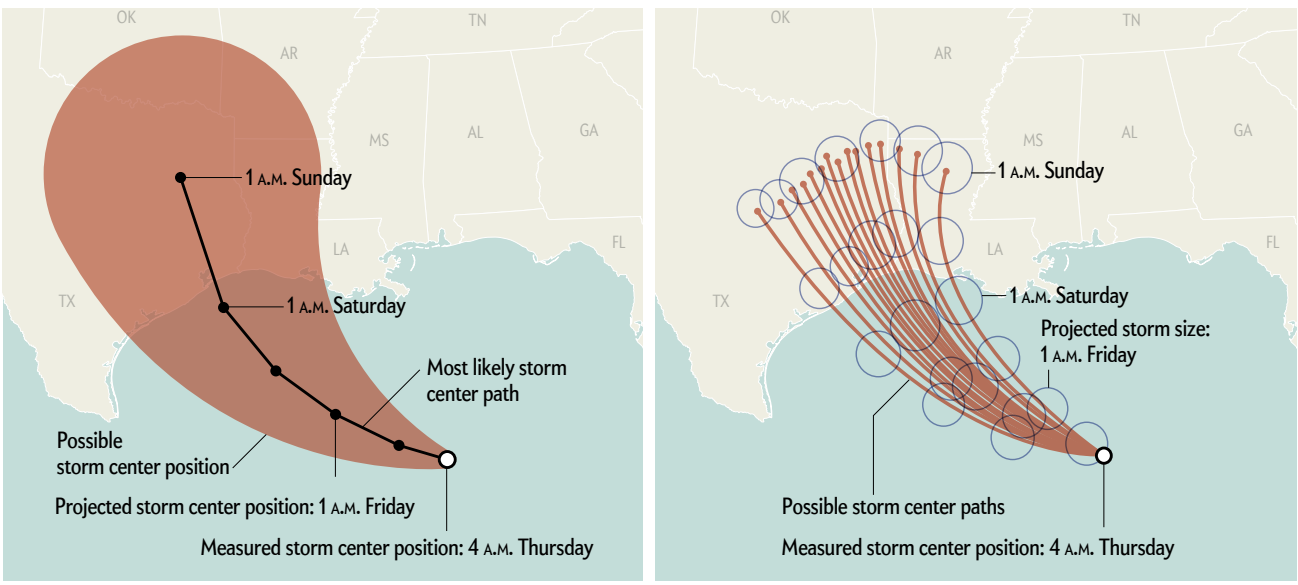
WHEN TRACKING A HURRICANE, FORECASTERS often show a map depicting a "cone of uncertainty." It starts as a point—the hurricane's current position—and widens into a swath of territory the storm might cross in the upcoming days. The most likely path is along the centerline of the cone, with the probability falling off toward the edges. The problem: many people misinterpret the cone as the size of the future storm.

Researchers have found that the misunderstanding can be prevented if forecasters instead show a number of possible paths. Yet this approach can also introduce misunderstanding: lots of people think the probability of damage is greater where each path intersects land and less likely between the lines (*maps*).

Uncertainty pervades the data that scientists and all kinds of organizations use to inform decisions. Visual depictions of information can help clarify the uncertainty—or compound confusion. Ideally, visualizations help us make judgments, analytically and emotionally, about the probability of different outcomes. Abundant evidence on human reasoning suggests, however, that when people are asked to make judgments involving probability, they often discount uncertainty. As society increasingly relies on data, graphics designers are grappling with how best to show uncertainty clearly.

What follows is a gallery of visualization techniques for displaying uncertainty, organized roughly from less effective to more effective. Seeing how different approaches are chosen and implemented can help us become more savvy consumers of data and the uncertainty involved. SA

Jessica Hullman is a professor of computer science and journalism at Northwestern University. She and her research group develop and evaluate data-visualization and data-interaction techniques to enhance reasoning about uncertainty.

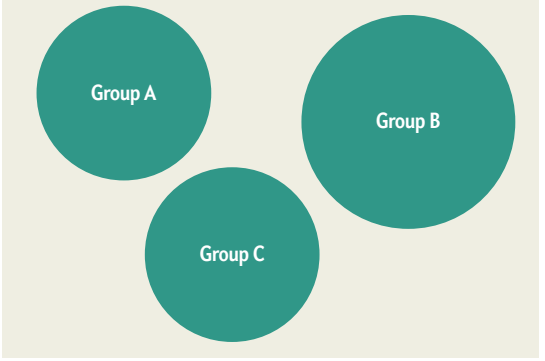


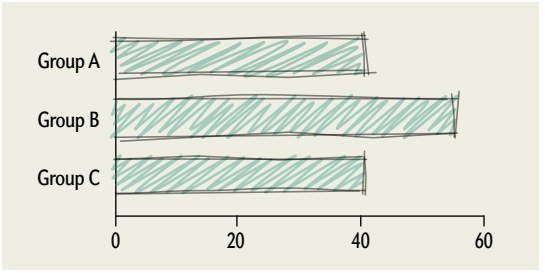
"CONE OF UNCERTAINTY" (*left*) shows where a hurricane may head, according to a group of forecasts. An alternative is to show the specific path predicted by each forecast (*right*). Both approaches have pros and cons in helping people judge the risk they may face, but the one on the right makes it clearer that the path is difficult to predict.

SOURCES: NATIONAL HURRICANE CENTER (*cone of uncertainty*); "VISUALIZING UNCERTAIN TROPICAL CYCLONE PREDICTIONS USING REPRESENTATIVE SAMPLES FROM ENSEMBLES OF FORECAST TRACKS," BY LE LIU ET AL., IN *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*, VOL. 25; AUGUST 20, 2018 (*multiple storm paths*)

Maps by Tiffany Farrant-Gonzalez

© 2020 Scientific American





NO QUANTIFICATION

The least effective way to present uncertainty is to not show it at all. Sometimes data designers try to compensate for a lack of specified uncertainty by choosing a technique that implies a level of imprecision but does not quantify it. For example, a designer might map data to a visual variable that is hard for people to define, such as a circle floating in space (*top*) rather than a dot on a graph that has *x* and *y* axes. This approach makes the reader's interpretation more error-prone. Alternatively a designer might use a program that creates a hand-drawn or "sketchy" feel (*bottom*). Both approaches are risky.

PROS

- If readers sense that a visualization is difficult to quantify or is simply impressionistic, they may be more cautious in making inferences or decisions based on it.

CONS

- Readers may not realize that the visualization is intended to convey imprecision and may reach conclusions that have large errors.
- Even if readers recognize that the visualization was chosen to imply imprecision, they have no way of inferring how much uncertainty is involved.

INTERVALS

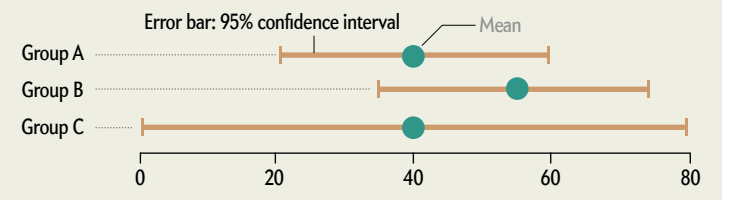
Intervals may be the most common representations of quantified uncertainty. Error bars (*top*) and confidence envelopes (*bottom*) are widely recognized, but even though they seem exact and straightforward, they are notoriously hard to interpret properly. Research shows they are often misunderstood, even by scientists.

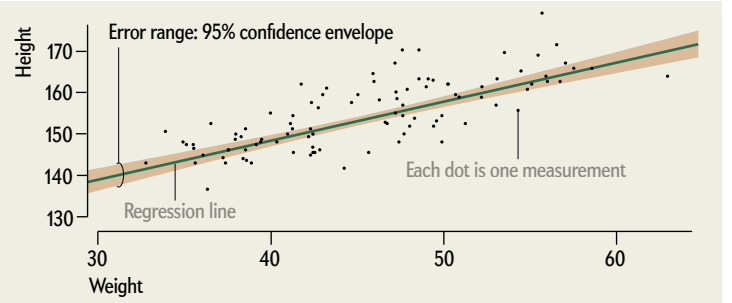
PROS

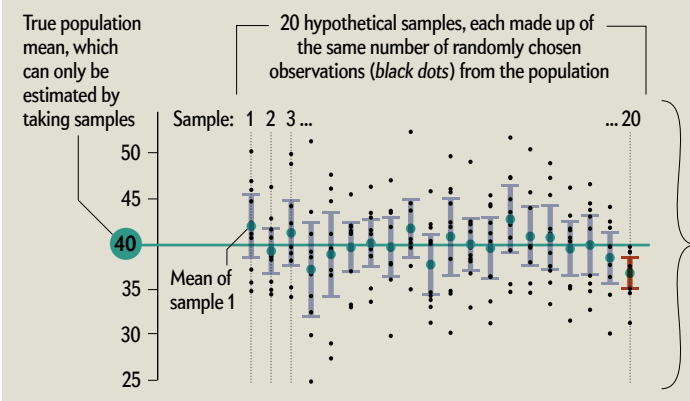
- Widely recognized as a representation of uncertainty.
- Offers a simple format for expressing the possibility of different values.
- The choice of interval can be customized for different types of questions about the same data set. For example, when one is making inferences about the range of values in a population, intervals based on standard deviation are helpful; for inferences about the range of values of a statistic like a mean, intervals based on standard error are appropriate.

CONS

- Ambiguity in what is shown: intervals may represent standard deviation, standard error or something else. Each has a unique interpretation.
- Readers can make "deterministic construal errors"—interpreting the ends of the error bar as the high and low values in observed measurements rather than estimates denoting uncertainty.
- Error bars can lead to "within-the-bar bias," common in bar charts. Below, readers may see the bar values to the right of the dots as more probable than the bar values to the left.
- Easy to ignore the uncertainty regions in favor of the central tendency, which may lead to incorrect decisions.






True population mean, which can only be estimated by taking samples
20 hypothetical samples, each made up of the same number of randomly chosen observations (*black dots*) from the population
Sample: 1 2 3 ...
... 20
50
45
40
35
30
25
Mean of sample 1

WHAT DOES A CONFIDENCE INTERVAL MEAN?

A natural interpretation of an error bar or confidence envelope that denotes 95 percent confidence is that the interval has a 95 percent chance of containing the true value. Yet it actually refers to the percentage of confidence intervals that would include the true value if an infinite number of random samples of the same size were pulled from the data and each time a 95 percent confidence interval was constructed. Although in practice this pervasive misinterpretation may not drastically change decisions, the fact that even scientists make such mistakes shows how challenging it can be to interpret uncertainty depictions correctly.

Even when calculated perfectly, on average, 1 in 20 of the 95% confidence intervals will not contain the population mean.

Graphics by Jessica Hullman and Jen Christiansen

© 2020 Scientific American

PROBABILITY DENSITY MAPS

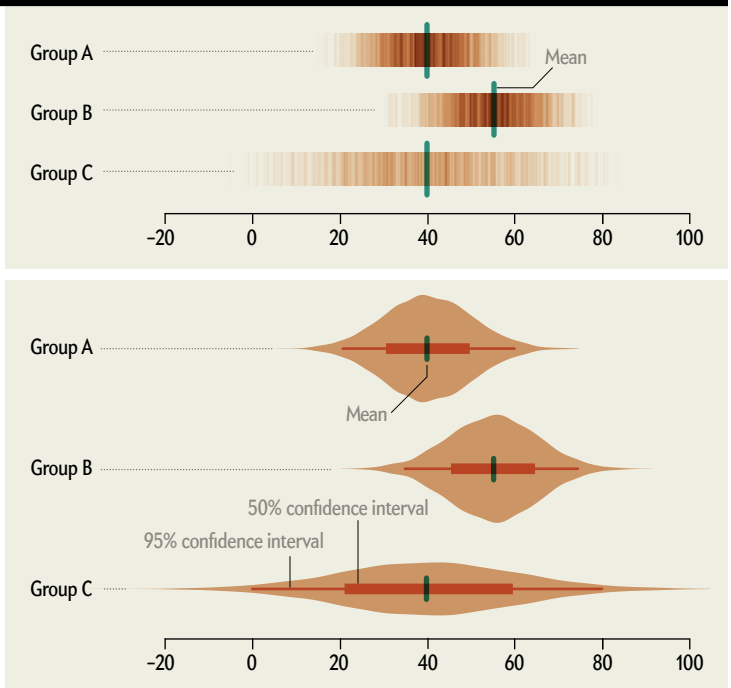
Designers can map uncertainty directly to a visual property of the visualization. For example, a gradient plot (*top*) can shift from dark color (high probability) at the center to lighter color (low probability) at the edges. In a violin plot (*bottom*), wider points mean greater probability. Mapping probability density to a visual variable displays uncertainty in greater detail than interval methods (error bars and confidence envelopes), but its effectiveness depends on how well readers can perceive differences in shading, height or other visual properties.

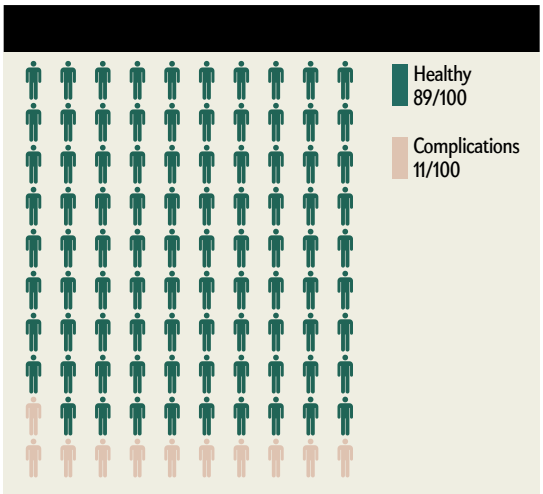
PROS

- Often well aligned with intuition: dark shading or hard boundaries are certain; light shading or fuzzy boundaries are uncertain.
- Avoids common biases such as those raised by intervals.

CONS

- Readers may not recognize that density reflects probability.
- Readers often equate the part of the visualization that is easiest to read (darkest, widest) with the data values themselves and misinterpret the parts that are harder to read (lightest, most narrow) as the uncertainty.
- Estimates can be biased to the darkest or highest points.
- Can be difficult to infer specific probability values.





ARRAYS OF ICONS

Reframing a probability such as 30 percent as a frequency—three out of 10—can make it easier for people to understand uncertainty and consequently use such information appropriately. People may better understand discrete probabilities because they run into them in everyday experiences.

PROS

- More self-explanatory than some other techaniques because readers can readily see that probability is analogous to the number of times a symbol appears.
- Readers can make quick estimates if a small number of symbols is used because our visual system recognizes small quantities immediately without counting.

CONS

- Designed to present only a single probability.

MULTIPLE SAMPLES IN SPACE

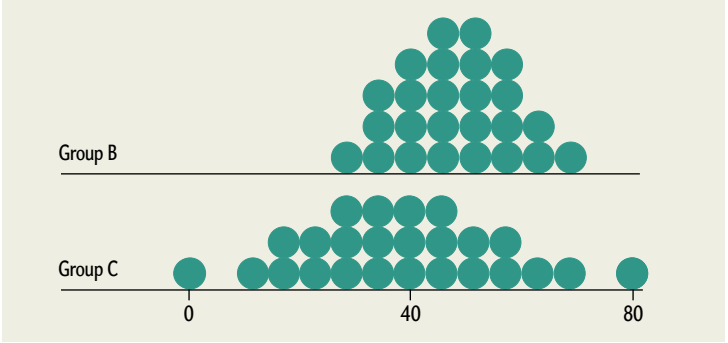
Plotting of multiple samples in space can be used to show probability in a discrete format for one or more variable quantities. One example of this approach is a quantile dot plot. It shows a number of distinct cases from the quantiles of the data distribution, so that the number of dots (such as two dots high or five dots high, in the example below) conveys probability. When there is uncertainty about parameter values from which estimates are drawn, such as initial conditions, samples can be generated that vary these parameters and can be shown in a single visualization.

PROS

- A designer can choose how many data samples to present, aiming to show enough to convey the distribution but not so many that it becomes difficult for a reader to make out the individual samples.

CONS

- Plotting many data samples can result in occlusion, making probability estimates more error-prone.
- Sampling introduces imprecision, especially if the underlying distribution is heavily skewed by outliers.



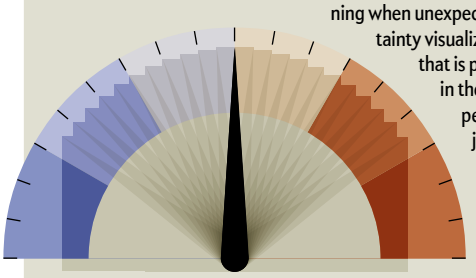
Graphics by Jessica Hullman and Jen Christiansen

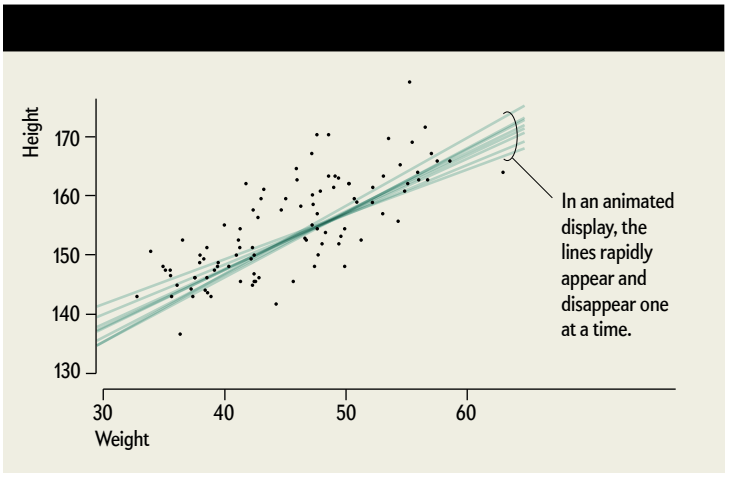
© 2020 Scientific American

A JITTERY ELECTION NIGHT NEEDLE

Sometimes an uncertainty visualization is controversial. On the night of the 2016 presidential election in the U.S., the *New York Times* introduced an animated gauge on its Web site to display predictions about the outcome. A continuum of colored areas made up the background, from a landslide Clinton win (*left*) to a landslide Trump win (*right*). The data model behind the gauge updated several times a minute as new local results came in. An animated needle jiggled back and forth rapidly, even more frequently than the model was updated.

Seeing a constantly moving visualization made many viewers anxious on an evening when unexpected events transpired. Uncertainty visualizations *should* provoke anxiety that is proportional to the uncertainty in the data. But after decades of people seeing only static projections for election outcomes that allowed viewers to overlook uncertainty, suddenly shifting to a visualization that provoked a visceral reaction to uncertainty was unsettling.





MULTIPLE SAMPLES IN TIME

Plotting multiple possible outcomes as frames in an animation makes uncertainty visceral and much harder to ignore. This technique, called hypothetical outcome plots, can be used for simple and complex visualizations. Perceptual studies indicate that people are surprisingly adept at inferring the distribution of data from the frequency of occurrences: we do not necessarily need to count the number of times an event occurs to estimate its probability. One important factor is the speed of events, which must be fast enough so that people can see a sufficient number of samples yet slow enough for them to consciously register what they saw.

PROS

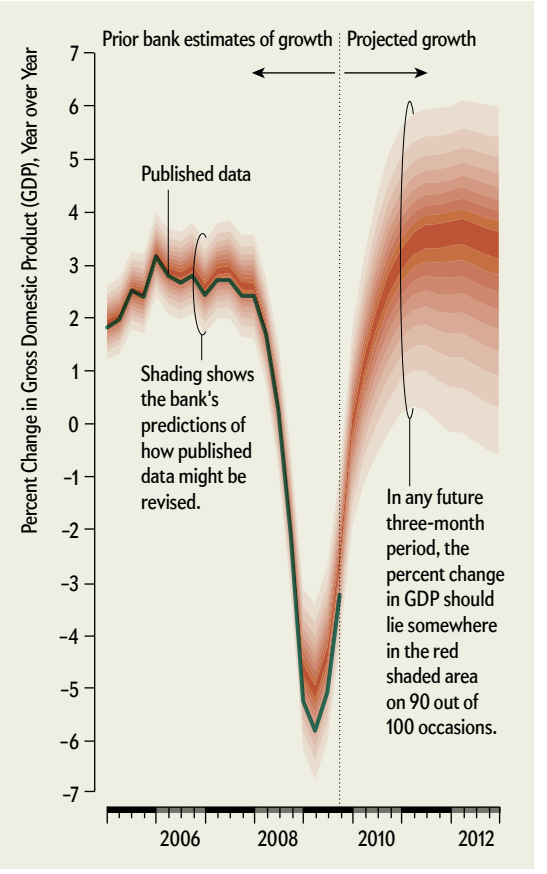
- The human visual system can estimate probability fairly accurately without having to deliberately count the items presented.
- Can be applied widely across different data types and visualization styles.
- Animation makes it possible to estimate probabilities involving multiple variables, which is difficult with static plots.

CONS

- Sampling introduces imprecision, especially if the distribution is heavily skewed by outliers.
- No guarantees on how many individual samples a user will pay attention to.
- Requires creating a dynamic or animated visualization, which some formats such as scientific papers may not yet easily support.

HYBRID APPROACHES

Designers can create effective uncertainty visualizations by combining different techniques rather than choosing a standard chart "type." One example is a fan chart, made famous by the Bank of England (*shown*). It depicts data up to the present (*left side of dotted line*), then projections into the future (*right side*); uncertainty about the past is an important component in assessing uncertainty about the future. The fan chart presents probability from higher chance (*dark shading*) to lesser chance (*light shading*) in multiple bands that represent different levels of confidence, which the reader can choose from. Readers can perceive the information through the position of the edges of the bands, as well as lightness versus darkness. Some modern software packages for statistical graphics and modeling make it easy to combine uncertainty visualization approaches.



SOURCE: *INFLATION REPORT*. BANK OF ENGLAND. FEBRUARY 2010 (*GDP chart*)

MORE TO EXPLORE

Picturing the Uncertain World: How to Understand, Communicate, and Control Uncertainty through Graphical Display. Harold Wainer. Princeton University Press, 2009.

Visualizing Uncertainty. Claus O. Wilke in *Fundamentals of Data Visualization*. O'Reilly Media, 2019.

Uncertainty + Visualization, Explained. Blog series by Jessica Hullman and Matthew Kay. https://medium.com/multiple-views-visualization-research-explained

scientificamerican.com/magazine/sa

Illustration by Tiffany Farrant-Gonzalez (election needle)

© 2020 Scientific American

Which Experts Should You Listen to during the Pandemic?

It should be a no-brainer: your best bet is to follow those who have actual expertise

By Nathan Ballantyne and David Dunning

As COVID-19 cases and fatalities surge around the world, a parallel surge of predictions and recommendations are competing for our attention. Commentators race to weigh in on this pandemic moment, eager to satisfy a pressing human need: facing an epidemic of both enormous consequence and enormous uncertainty, we all crave answers. We want to keep ourselves and loved ones safe. We want the assurance of knowledge.

But which commentators should we trust? There are many voices, but they often fail to harmonize. Some even sing dangerous tunes. The cacophony requires us to weigh carefully which voices we should heed.

First and foremost, we should filter out those who are addressing issues beyond their intellectual training and competence. In philosophy, this phenomenon has a name: epistemic trespassing. It occurs when commentators holding real expertise in one field intrude into another, passing judgment where they lack crucial competence. Roaming into a field without expert-level insight, trespassers easily slip up.

We have witnessed a cavalcade of epistemic trespassing on pandemic topics. In mid-March a prominent American legal scholar argued that official estimates for COVID-19 deaths in the U.S. were radically overinflated. His theoretical model indicated that "about 500" Americans would die. The head of the Israel Space Agency, a military scientist, produced a statistical model predicting that the virus would peak after 40 days and fall to nearly zero by 70, no matter what containment measures are used. Well beyond that 70-day limit in the U.S., zero is nowhere in sight.

Then there is medical treatment. An unlicensed dermatologist from Portland, Ore., promoted "medical 'chicken pox' parties" where people would voluntarily infect themselves with coronavirus. In Iran, some religious leaders have instructed people to use medicinal herbs, oils and fruit juices to guard against the virus.

This sort of trespassing is nothing new. John Barry's *The Great Influenza,* a book about the horrors of the 1918 pandemic, describes people stocking up on fraudulent elixirs and hanging garlic around their necks. Newspapers ran advertisements that were hard to distinguish from real articles. A shoe store advertised, "One way to keep the flu away is to keep your feet dry."

Trespassing by faux authority historically has spread across many scientific fields. One troubling example, described in detail by Naomi Oreskes and Erik M. Conway in their book *Merchants of Doubt,* is an elite cadre of 20th-century American scientists who spread misguided messages over the course of several decades on global warming, acid rain and the health risks of smoking cigarettes. Bankrolled by industries with significant financial interests, they confused both public and politicians about what the actual science had established.

Why do people trespass, despite the risks to others and to their own reputations? We presume (appealing to Hanlon's razor: "Never attribute to malice that which can be adequately explained by stupidity") that most have honorable motives. But good judgment requires more than just good intentions.

Getting traction in any domain of knowledge—virology, medical treatment and public health in this case—calls for specific skills and knowledge to interpret evidence accu-

© 2020 Scientific American

rately and make recommendations. Genuine experts have capacities precisely tailored to the problems they tackle, allowing them to reach sound conclusions or at least reduce mistakes. They bring years of relevant training, historical knowledge and an appreciation of their limits. Of course, even when experts stick to their home fields, they are not infallible, especially when confronting novel challenges. But through their skill, true experts are far better at recognizing uncertainties and adjusting their judgments in light of new evidence.

SNAKE OIL SALESMEN knowingly duped customers with fraudulent cures. "Epistemic trespassers" often erroneously believe that their skill set carries to other fields.

GETTY IMAGES

Trespassers lack that well-tailored expertise. What they actually do know does not always transfer to new and different topics. Worse, they often lack the awareness that such tailored expertise exists. Their gaps in knowledge remain invisible to them.

Trespassers believe instead that their admittedly valuable "home field" skills carry over to other fields. Knowing only those skills, they believe they need nothing more to step outside their true circle of competence. They are like the proverbial worker who knows how to use a hammer and thus sees everything as a nail—not fathoming the existence of other tools.

To be sure, newcomers occasionally bring valuable insights to other established fields. In 1915, for example, German scientist Alfred Wegener, an applied astronomer and meteorologist, famously proposed his theory of continental drift. The vast majority of card-carrying geologists dismissed his ideas, but today core notions of his theory are taught in geology classrooms under the updated name "plate tectonics."

But one general lesson for outsiders is that the best recipe for making an impact is through collaboration with experts trained in the relevant discipline.

Coronavirus trespassers strike out alone on social media and in op-ed articles. They may crowd out real experts and science communicators trying to mitigate misinformation. One infectious disease epidemiologist recently complained of "an epidemic of armchair epidemiology." As a commentator once put it, the proliferation of spurious claims is like "the releasing of dummy hares for serious researchers to waste their time and energy chasing." It is costly for experts to catch and correct trespassers' errors, and success is uncertain.

While the rest of us navigate hectic, confusing channels of information, the concept of epistemic trespassing can help us decide whom to trust. One key insight is that judging another person's expertise often requires us to trespass beyond our own circle of competence. So how do we choose which commentators to follow and which to sideline?

Here are three essential questions for trusting more wisely. First, what fields of knowledge are strictly relevant for determining the message's accuracy? Is the issue a matter for epidemiologists, immunologists or vaccine science?

Second, does the messenger have training and a track record of success, in that relevant field? We nonexperts cannot always expect to distinguish trespassers from genuine experts, but our failure is more certain if we do not pause to ask sharp questions about credibility. Some trespassers may have impressive street cred, such as success in business or an Ivy League degree. Our advice is to not be impressed by these general haloes but instead to focus on specific credentials.

Third, and most important, ask whether multiple and independent sources concur. Do other messengers state the same claim? Is the messenger speaking on behalf of many experts? It is essential to cross-check assertions by those who claim expert status. Trust not any single expert, but experts in the plural.

In anxious times, we all feel the need to know. But as we pursue satisfying that need, we must be cautious. Even the most highly educated and successful among us are not immune to gullibility. Humility about the limits of our own knowledge is imperative, including when trying to identify true expertise in others. Seeking knowledge—now as much as ever—means keeping an open mind, checking sources conscientiously and sometimes admitting "I don't know." Being prepared to make that last confession may be the surest way to ensure our future arrival at the truth. SA

Nathan Ballantyne is an associate professor of philosophy at Fordham University and is author of *Knowing Our Limits* (Oxford University Press, 2019).

David Dunning is a professor of psychology at the University of Michigan, most known for his work on the Dunning-Kruger effect and other mistakes of self-judgment. He co-authored with Kaidi Wu the essay *Unknown Unknowns: The Problem of Hypocognition* for ScientificAmerican.com in 2018.

MORE TO EXPLORE

The Dunning–Kruger Effect: On Being Ignorant of One's Own Ignorance. David Dunning in *Advances in Experimental Social Psychology,* Vol. 44, pages 247-296; 2011.

scientificamerican.com/magazine/sa

© 2020 Scientific American

Big Data and Small Decisions

For individuals a deluge of facts can be a problem

By Zeynep Tufekci

In 2018, as back-to-back hurricanes Florence and Michael threatened Chapel Hill, N.C., where I live and work, I faced a simple, binary decision like millions of others: Stay or go?

Nowadays data science is the hottest thing around. Companies cannot hire enough practitioners. There are books and online courses, and many universities are launching some flavor of a data science degree or center. Classes can barely accommodate the demand. One would hope that this golden age would mean we can make better decisions. But the deluge of data can, paradoxically, make decision-making harder: it doesn't easily translate into useful information. The democratization of access and the proliferation of expert commentary can make things even thornier. Finally, measurement itself is not a neutral process.

The days leading up to the landfall of both hurricanes, for example, were dominated by their number on the familiar Saffir-Simpson categories of 1 to 5, corresponding to sustained wind speeds, along with the "cone" of the storms' probable trajectories. Outside mandatory evacuation zones, it was up to everyone to decide for themselves what to do. As management consultant Peter Drucker is credited with saying: "If you can't measure it, you can't improve it." I'd add: "If you *do* measure it, you'll be trapped by the number." That's the problem with wind intensities: wind damage is obviously relevant, but the worst impact can come from flooding. Florence came ashore as a mere category1, then dumped *three feet* of rain in some places—including away from the cone.

Seeking clarity, I checked in on the local TV meteorologists, who could pinpoint local impacts beyond one number. But arguably they had a bias toward emphasizing the dangers, which is better both for ratings and for self-preservation: it's much dicier if people don't evacuate when they should than if they flee unnecessarily. So I geared up to find more data. I sought out weather experts on social media and found well-curated lists. It seemed like a great idea at first. These were genuine experts. The commentary was respectful and intelligent. There were links to sources, and the discussion was rich.

But I quickly remembered why I never want to watch sausage being made. I learned a lot about European versus North American weather models—fascinating but fairly useless when you're trying to decide whether to pack up a few sentimental photographs and leave. One model predicted devastation, the other just some heavy rain. A storm could turn north, for a direct hit, or south—a miss. Worse, each model updated periodically, each run generating more expert discussion.

Now I knew too much but had even less clarity for decision-making. This is sometimes referred to as the "paradox of choice"—too many options can paralyze people trying to make a decision. It's that feeling you get standing in front of the ketchup shelves in the supermarket, overwhelmed. Organic or not? Low sugar? Sweetened with honey? With artificial sweeteners—and if so, sucralose or aspartame? Low sodium? I have resorted to blindly grabbing one—I just want a bottle of ketchup. (Well, glass or plastic?)

So if more data, better science and mightier computation can give us a hurricane's trajectory so many days in advance, why can't anyone make better predictions of impacts at the hyperlocal level? Unfortunately, broad predictions don't easily trickle down, because individual outcomes retain big error ranges—too many false positives and false negatives to be easily actionable.

So should we give up on data-driven decision-making in our own lives? Like many things in the age of big data, the way forward requires paying attention to things beyond the data—from how and what to measure to how to communicate about it. We need more frank talk about the shortcomings, so we can refine our understanding of the difference between a lot of data and useful information. And we especially need to build independent intermediaries to help guide us. Data science by itself can't do all that.

As for the hurricanes, I had just moved to my street so I did the simplest thing I could think of: I asked my neighbors who'd been there for a long time. They advised me to stock up on batteries. They stayed put, and so did I. SA

Zeynep Tufekci is an associate professor at the University of North Carolina, whose research revolves around how technology, science and society interact.

Illustration by Sam Island

© 2020 Scientific American

When "Like" Is a Weapon

Everyone is an agent in the new information warfare

By the Editors

NO ONE THINKS, I AM THE KIND OF PERSON WHO is susceptible to misinformation. It is those others (stupid anti-vaxxers! arrogant liberal elites!) who are swayed by propaganda masquerading as news and bot armies pushing partisan agendas on Twitter.

But recent disinformation campaigns—especially ones that originate with coordinated agencies in Russia or China—have been far more sweeping and insidious. Using memes, manipulated videos and impersonations to spark outrage and confusion, their targets transcend any single election or community. Indeed, these efforts aim to engineer volatility to undermine democracy itself. If we're all mentally exhausted and we disagree about what is true, then authoritarian networks can more effectively push their version of reality. Playing into the "us versus them" dynamic makes everyone more vulnerable to false belief.

Instead of surrendering to the idea of a post-truth world, we must recognize this so-called information disorder as an urgent societal crisis and bring rigorous, interdisciplinary scientific research to combat the problem. We need to understand the transmission of knowledge online; the origins, motivations and tactics of disinformation networks, both foreign and domestic; and exactly how even the most educated evidence seekers can unwittingly become part of an influence operation. Little is known, for instance, about the effects of long-term exposure to disinformation or how it affects our brain or voting behavior. To examine these connections, technology behemoths such as Facebook, Twitter and Google must make more of their data available to independent researchers (while protecting user privacy).

The pace of research must try to catch up with the rapidly growing sophistication of disinformation strategies. One positive step was the launch in January of the *Misinformation Review,* a multimedia-format journal from Harvard University's John F. Kennedy School of Government that will fast-track its peer-review process and prioritize articles about real-world implications of misinformation in areas such as the media, public health and elections.

Journalists must be trained in how to cover deception so that they don't inadvertently entrench it, and governments should strengthen their information agencies to fight back. Western nations can look to the Baltic states to learn some of the innovative ways their citizens have dealt with disinformation over the past decade: for example, volunteer armies of civilian "elves" expose the methods of Kremlin "trolls." Minority and historically oppressed communities are also familiar with ways to push back on authorities' attempts to overwrite truth. Critically, technologists should collaborate with social scientists to propose interventions—and they would be wise to imagine how attackers might thwart these tools or turn them around to use for their own means.

Ultimately, though, for most disinformation operations to succeed, it is regular users of the social Web who must share the videos, use the hashtags and add to the inflammatory comment threads. That means each one of us is a node on the battlefield for reality. Individuals need to be more aware of how our emotions and biases can be exploited with precision and consider what forces might be provoking us to amplify divisive messages.

So every time you want to "like" or share a piece of content, imagine a tiny "pause" button hovering over the thumbs-up icon on Facebook or the retweet symbol on Twitter. Hit it and ask yourself, Am I responding to a meme meant to brand me as a partisan on a given issue? Have I actually read the article, or am I simply reacting to an amusing or enraging headline? Am I sharing this piece of information only to display my identity for my audience of friends and peers, to get validation through likes? If so, what groups might be microtargeting *me* through my consumer data, political preferences and past behavior to manipulate me with content that resonates strongly?

Even if—especially if—you're passionately aligned with or disgusted by the premise of a meme, ask yourself if sharing it is worth the risk of becoming a messenger for disinformation meant to divide people who might otherwise have much in common.

It is easy to assume that memes are innocuous entertainment, not powerful narrative weapons in a battle between democracy and authoritarianism. But these are among the tools of the new global information wars, and they will only evolve as machine learning advances. If researchers can figure out what would get people to take a reflective pause, it may be one of the most effective ways to safeguard public discourse and reclaim freedom of thought. SA

Illustration by Aad Goudappel

© 2020 Scientific American

A New World Disorder

Our willingness to share content without thinking is exploited to spread disinformation

By Claire Wardle

Illustration by Wesley Allsbrook

© 2020 Scientific American

© 2020 Scientific American

As someone who studies the impact of misinformation on society, I often wish the young entrepreneurs of Silicon Valley who enabled communication at speed had been forced to run a 9/11 scenario with their technologies before they deployed them commercially.

One of the most iconic images from that day shows a large clustering of New Yorkers staring upward. The power of the photograph is that we know the horror they're witnessing. It is easy to imagine that, today, almost everyone in that scene would be holding a smartphone. Some would be filming their observations and posting them to Twitter and Facebook. Powered by social media, rumors and misinformation would be rampant. Hate-filled posts aimed at the Muslim community would proliferate, the speculation and outrage boosted by algorithms responding to unprecedented levels of shares, comments and likes. Foreign agents of disinformation would amplify the division, driving wedges between communities and sowing chaos. Meanwhile those stranded on the tops of the towers would be livestreaming their final moments.

Stress-testing technology in the context of the worst moments in history might have illuminated what social scientists and propagandists have long known: that humans are wired to respond to emotional triggers and share misinformation if it reinforces existing beliefs and prejudices. Instead designers of the social platforms fervently believed that connection would drive tolerance and counteract hate. They failed to see how technology would not change who we are fundamentally—it could only map onto existing human characteristics.

Online misinformation has been around since the mid-1990s. But in 2016 several events made it broadly clear that darker forces had emerged: automation, microtargeting and coordination were fueling information campaigns designed to manipulate public opinion at scale. Journalists in the Philippines started raising flags as Rodrigo Duterte rose to power, buoyed by intensive Facebook activity. This was followed by unexpected results in the Brexit referendum in June and then the U.S. presidential election in November—all of which sparked researchers to systematically investigate the ways in which information was being used as a weapon.

During the past four years the discussion around the causes of our polluted information ecosystem has focused almost entirely on actions taken (or not taken) by the technology companies. But this fixation is too simplistic. A complex web of societal shifts is making people more susceptible to misinformation and conspiracy. Trust in institutions is falling because of political and economic upheaval, most notably through ever widening income inequality. The effects of climate change are becoming more pronounced. Global migration trends spark concern that communities will change irrevocably. The rise of automation makes people fear for their jobs and their privacy.

Bad actors who want to deepen existing tensions understand these societal trends, designing content that they hope will so anger or excite targeted users that the audience will become the messenger. The goal is that users will use their own social capital to reinforce and give credibility to that original message.

Most of this content is designed not to persuade people in any particular direction but to cause confusion, to overwhelm and to undermine trust in democratic institutions from the electoral system to journalism. And although much is being made about preparing the U.S. electorate for the 2020 election, misleading and conspiratorial content did not begin with the 2016 presidential race, and it will not end after this one. As tools designed to manipulate and amplify content become cheaper and more accessible, it will be even easier to weaponize users as unwitting agents of disinformation.

IN BRIEF

Many types of information disorder exist online, from fabricated videos to impersonated accounts to memes designed to manipulate genuine content.
Automation and microtargeting tactics have made it easier for agents of disinformation to weaponize regular users of the social web to spread harmful messages.
Much research is needed to understand the effects of disinformation and build safeguards against it.

© 2020 Scientific American

WEAPONIZING CONTEXT

GENERALLY, THE LANGUAGE used to discuss the misinformation problem is too simplistic. Effective research and interventions require clear definitions, yet many people use the problematic phrase "fake news." Used by politicians around the world to attack a free press, the term is dangerous. Recent research shows that audiences increasingly connect it with the mainstream media. It is often used as a catchall to describe things that are not the same, including lies, rumors, hoaxes, misinformation, conspiracies and propaganda, but it also papers over nuance and complexity. Much of this content does not even masquerade as news—it appears as memes, videos and social posts on Facebook and Instagram.

In February 2017 I created seven types of "information disorder" in an attempt to emphasize the spectrum of content being used to pollute the information ecosystem. They included, among others, satire, which is not intended to cause harm but still has the potential to fool; fabricated content, which is 100 percent false and designed to deceive and do harm; and false context, which is when genuine content is shared with false contextual information. Later that year technology journalist Hossein Derakhshan and I published a report that mapped out the differentiations among disinformation, misinformation and malinformation.

Purveyors of *dis*information—content that is intentionally false and designed to cause harm—are motivated by three distinct goals: to make money; to have political influence, either foreign or domestic; and to cause trouble for the sake of it.

Those who spread *mis*information—false content shared by a person who does not realize it is false or misleading—are driven by sociopsychological factors. People are performing their identities on social platforms to feel connected to others, whether the "others" are a political party, parents who do not vaccinate their children, activists who are concerned about climate change, or those who belong to a certain religion, race or ethnic group. Crucially, disinformation can turn into misinformation when people share disinformation without realizing it is false.

We added the term "*mal*information" to describe genuine information that is shared with an intent to cause harm. An example of this is when Russian agents hacked into e-mails from the Democratic National Committee and the Hillary Clinton campaign and leaked certain details to the public to damage reputations.

Having monitored misinformation in eight elections around the world since 2016, I have observed a shift in tactics and techniques. The most effective disinformation has always been that which has a kernel of truth to it, and indeed most of the content being disseminated now is not fake—it is misleading. Instead of wholly fabricated stories, influence agents are reframing genuine content and using hyperbolic headlines. The strategy involves connecting genuine content with polarizing topics or people. Because bad actors are always one step (or many steps) ahead of platform moderation, they are relabeling emotive disinformation as satire so that it will not get picked up by fact-checking processes. In these efforts, context, rather than content, is being weaponized. The result is intentional chaos.

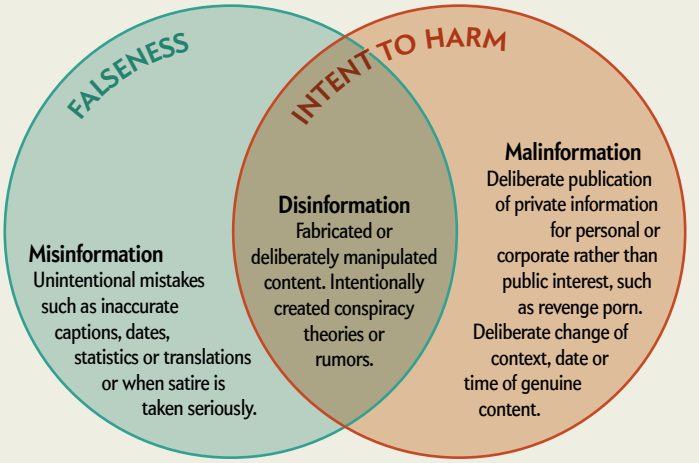
Take, for example, the edited video of House Speaker Nancy Pelosi that circulated in May 2019. It was a genuine video, but an agent of disinformation slowed down the video and then posted that clip to make it seem that Pelosi was slurring her words. Just as intended, some viewers immediately began speculating that Pelosi was drunk, and the video spread on social media. Then the mainstream media picked it up, which undoubtedly made many more people aware of the video than would have originally encountered it.

Research has found that traditionally reporting on misleading content can potentially cause more harm. Our brains are wired to rely on heuristics, or mental shortcuts, to help us judge credibility. As a result, repetition and familiarity are two of the most effective mechanisms for ingraining misleading narratives, even when viewers have received contextual information explaining why they should know a narrative is not true.

Bad actors know this: In 2018 media scholar Whitney Phillips published a report for the Data & Society Research Institute that explores how those attempting to push false and misleading narratives use techniques to encourage reporters to cover their narratives. Yet another report from the Institute for the Future found that only 15 percent of U.S. journalists had been trained in how to report on misinformation more responsibly. A central challenge now for reporters and fact check-

THREE CATEGORIES OF INFORMATION DISORDER

To understand and study the complexity of the information ecosystem, we need a common language. The current reliance on simplistic terms such as "fake news" hides important distinctions and denigrates journalism. It also focuses too much on "true" versus "fake," whereas information disorder comes in many shades of "misleading."


FALSENESS
INTENT TO HARM
Misinformation
Unintentional mistakes such as inaccurate captions, dates, statistics or translations or when satire is taken seriously.
Disinformation
Fabricated or deliberately manipulated content. Intentionally created conspiracy theories or rumors.
Malinformation
Deliberate publication of private information for personal or corporate rather than public interest, such as revenge porn. Deliberate change of context, date or time of genuine content.

GRAPHIC BY JEN CHRISTIANSEN; SOURCE: *INFORMATION DISORDER: TOWARD AN INTERDISCIPLINARY FRAMEWORK FOR RESEARCH AND POLICYMAKING*, BY CLAIRE WARDLE AND HOSSEIN DERAKHSHAN, COUNCIL OF EUROPE, OCTOBER 2017

© 2020 Scientific American

ers—and anyone with substantial reach, such as politicians and influencers—is how to untangle and debunk falsehoods such as the Pelosi video without giving the initial piece of content more oxygen.

MEMES: A MISINFORMATION POWERHOUSE

IN JANUARY 2017 the NPR radio show *This American Life* interviewed a handful of Trump supporters at one of his inaugural events called the DeploraBall. These people had been heavily involved in using social media to advocate for the president. Of Trump's surprising ascendance, one of the interviewees explained: "We memed him into power.... We directed the culture."

The word "meme" was first used by theorist Richard Dawkins in his 1976 book, *The Selfish Gene,* to describe "a unit of cultural transmission or a unit of imitation," an idea, behavior or style that spreads quickly throughout a culture. During the past several decades the word has been appropriated to describe a type of online content that is usually visual and takes on a particular aesthetic design, combining colorful, striking images with block text. It often refers to other cultural and media events, sometimes explicitly but mostly implicitly.

OF TRUMP'S SURPRISING ASCENDANCE, ONE OF THE DEPLORABALL INTERVIEWEES EXPLAINED: "WE MEMED HIM INTO POWER.... WE DIRECTED THE CULTURE."

This characteristic of implicit logic—a nod and wink to shared knowledge about an event or person—is what makes memes impactful. Enthymemes are rhetorical devices where the argument is made through the absence of the premise or conclusion. Often key references (a recent news event, a statement by a political figure, an advertising campaign or a wider cultural trend) are not spelled out, forcing the viewer to connect the dots. This extra work required of the viewer is a persuasive technique because it pulls an individual into the feeling of being connected to others. If the meme is poking fun or invoking outrage at the expense of another group, those associations are reinforced even further.

The seemingly playful nature of these visual formats means that memes have not been acknowledged by much of the research and policy community as influential vehicles for disinformation, conspiracy or hate. Yet the most effective misinformation is that which will be shared, and memes tend to be much more shareable than text. The entire narrative is visible in your feed; there is no need to click on a link. A 2019 book by An Xiao Mina, *Memes to Movements,* outlines how memes are changing social protests and power dynamics, but this type of serious examination is relatively rare.

Indeed, of the Russian-created posts and ads on Facebook related to the 2016 election, many were memes. They focused on polarizing candidates such as Bernie Sanders, Hillary Clinton or Donald Trump and on polarizing policies such as gun rights and immigration. Russian efforts often targeted groups based on race or religion, such as Black Lives Matter or Evangelical Christians. When the Facebook archive of Russian-generated memes was released, some of the commentary at the time centered on the lack of sophistication of the memes and their impact. But research has shown that when people are fearful, oversimplified narratives, conspiratorial explanation, and messages that demonize others become far more effective. These memes did just enough to drive people to click the share button.

Technology platforms such as Facebook, Instagram, Twitter and Pinterest play a significant role in encouraging this human behavior because they are designed to be performative in nature. Slowing down to check whether content is true before sharing it is far less compelling than reinforcing to your "audience" on these platforms that you love or hate a certain policy. The business model for so many of these platforms is attached to this identity performance because it encourages you to spend more time on their sites.

Researchers are now building monitoring technologies to track memes across different social platforms. But they can investigate only what they can access, and the data from visual posts on many social platforms are not made available to researchers. Additionally, techniques for studying text such as natural-language processing are far more advanced than techniques for studying images or videos. That means the research behind solutions being rolled out is disproportionately skewed toward text-based tweets, Web sites or articles published via URLs, and fact-checking of claims by politicians in speeches.

Although plenty of blame has been placed on the technology companies—and for legitimate reasons—they are also products of the commercial context in which they operate. No algorithmic tweak, update to the platforms' content-moderation guidelines or regulatory fine will alone improve our information ecosystem at the level required.

PARTICIPATING IN THE SOLUTION

IN A HEALTHY INFORMATION commons, people would still be free to express what they want—but information that is designed to mislead, incite hatred, reinforce tribalism or cause physical harm would not be amplified by algorithms. That means it would not be allowed

© 2020 Scientific American

to trend on Twitter or in the YouTube content recommender. Nor would it be chosen to appear in Facebook feeds, Reddit searches or top Google results.

Until this amplification problem is resolved, it is precisely our willingness to share without thinking that agents of disinformation will use as a weapon. Hence, a disordered information environment requires that every person recognize how he or she, too, can become a vector in the information wars and develop a set of skills to navigate communication online as well as offline.

Currently conversations about public awareness are often focused on media literacy and often with a paternalistic framing that the public simply needs to be taught how to be smarter consumers of information. Instead online users would be better taught to develop cognitive "muscles" in emotional skepticism and trained to withstand the onslaught of content designed to trigger base fears and prejudices.

Anyone who uses Web sites that facilitate social interaction would do well to learn how they work—and especially how algorithms determine what users see by "prioritiz[ing] posts that spark conversations and meaningful interactions between people," in the case of a January 2018 Facebook update about its rankings. I would also recommend that everyone try to buy an advertisement on Facebook at least once. The process of setting up a campaign helps to drive understanding of the granularity of information available. You can choose to target a subcategory of people as specific as women, aged between 32 and 42, who live in the Raleigh-Durham area of North Carolina, have preschoolers, have a graduate degree, are Jewish and like Kamala Harris. The company even permits you to test these ads in environments that allow you to fail privately. These "dark ads" let organizations target posts at certain people, but they do not sit on that organization's main page. This makes it difficult for researchers or journalists to track what posts are being targeted at different groups of people, which is particularly concerning during elections.

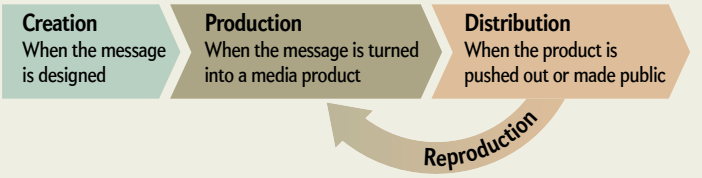
Facebook events are another conduit for manipulation. One of the most alarming examples of foreign interference in a U.S. election was a protest that took place in Houston, Tex., yet was entirely orchestrated by trolls based in Russia. They had set up two Facebook pages that looked authentically American. One was named "Heart of Texas" and supported secession; it created an "event" for May 21, 2016, labeled "Stop Islamification of Texas." The other page, "United Muslims of America," advertised its own protest, entitled "Save Islamic Knowledge," for the exact same time and location. The result was that two groups of people came out to protest each other, while the real creators of the protest celebrated the success at amplifying existing tensions in Houston.

Another popular tactic of disinformation agents is dubbed "astroturfing." The term was initially connected to people who wrote fake reviews for products online or tried to make it appear that a fan community was larger than it really was. Now automated campaigns use bots or the sophisticated coordination of passionate supporters and paid trolls, or a combination of both, to make it appear that a person or policy has considerable grassroots support. By making certain hashtags trend on Twitter, they hope that particular messaging will get picked up by the professional media and direct the amplification to bully specific people or organizations into silence.

Understanding how each one of us is subject to such campaigns—and might unwittingly participate in them—is a crucial first step to fighting back against those who seek to upend a sense of shared reality. Perhaps most important, though, accepting how vulnerable our society is to manufactured amplification needs to be done sensibly and calmly. Fearmongering will only fuel more conspiracy and continue to drive down trust in quality-information sources and institutions of democracy. There are no permanent solutions to weaponized narratives. Instead we need to adapt to this new normal. Just as putting on sunscreen was a habit that society developed over time and then adjusted as additional scientific research became available, building resiliency against a disordered information environment needs to be thought about in the same vein. SA

HOW DISINFORMATION BECOMES MISINFORMATION

The spread of false or misleading information is often dynamic. It starts when a disinformation agent engineers a message to cause maximum harm—for example, designing real-life protests that put opposing groups in public conflict. In the next phase, the agent creates "Event" pages on Facebook. The links are pushed out to communities that might be intrigued. People who see the event are unaware it is a false premise and share it with their communities, using their own framing. This reproduction continues.



GRAPHIC BY JEN CHRISTIANSEN; SOURCE: *INFORMATION DISORDER: TOWARD AN INTERDISCIPLINARY FRAMEWORK FOR RESEARCH AND POLICYMAKING*, BY CLAIRE WARDLE AND HOSSEIN DERAKHSHAN, COUNCIL OF EUROPE, OCTOBER 2017

Claire Wardle is U.S. director at the nonprofit First Draft, where she researches, leads projects and trains groups on understanding and tackling misinformation. Previously she was a research fellow at the Shorenstein Center on Media, Politics and Public Policy at Harvard University. She has a Ph.D. in communication from the University of Pennsylvania.

MORE TO EXPLORE

Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making. Claire Wardle and Hossein Derakhshan. Council of Europe, October 2017.
Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics. Yochai Benkler, Robert Faris and Hal Roberts. Oxford University Press, 2018.
Memes to Movements: How the World's Most Viral Media Is Changing Social Protest and Power. An Xiao Mina. Beacon Press, 2019.
Priming and Fake News: The Effects of Elite Discourse on Evaluations of News Media. Emily Van Duyn and Jessica Collier in *Mass Communication and Society,* Vol. 22, No. 1, pages 29–48; 2019.

scientificamerican.com/magazine/sa

© 2020 Scientific American

INSIDE THE ECHO CHAMBER

Data scientists are studying how information spreading online influences our social dynamics and what, if anything, can be done to smooth polarization

By Walter Quattrociocchi

IN BRIEF

Despite optimistic talk about "collective intelligence," the Web has helped create an echo chamber where misinformation thrives. Indeed, the viral spread of hoaxes, conspiracy theories, and other false or baseless information online is one of the most disturbing social trends of the early 21st century. **Data scientists** are studying this echo chamber by applying computational methods to the traces people leave on Facebook, Twitter and other such outlets. Through this work, they have established that users happily embrace false information as long as it reinforces their preexisting beliefs. **Faced with complex global issues,** people of all educational levels choose to believe compact—but false—explanations that clearly identify an object of blame. Unfortunately, attempts to debunk false beliefs seem only to reinforce polarization. Stopping the spread of misinformation is thus a problem with no apparent simple solution.

Illustration by Chris Malbon

© 2020 Scientific American

© 2020 Scientific American

IN THE SUMMER OF 2015 GOVERNOR Greg Abbott gave the Texas State Guard an unusual order: keep an eye on the Jade Helm 15 exercise, just in case the online rumors are true. In reality, Jade Helm 15 was a routine eight-week military exercise conducted in Texas and six other states. In the online echo chamber, however, it was something more sinister: the beginning of a coup ordered by President Barack Obama.

One great promise of the creation of the World Wide Web was that users might be exposed to diversified points of view. The human attention span, however, remains limited, and news feed algorithms may favor selective exposure. Users show the tendency to select information that adheres to their beliefs and join polarized groups formed around a shared narrative, called echo chambers. In these closed communities, polarization is dominant. In this environment, unreliable rumors and conspiracy theories spread widely. Conspiracy theories are nothing new, but in an age of rampant populism and digital activism, they have acquired new power to influence real-world events—usually for the worse. In a 2017 report on global risks, the World Economic Forum named the polarization and the viral spread of biased information as one of the most dangerous social trends of the age. With antidemocratic politicians on the rise throughout the West, we are now seeing the danger of viral misinformation become manifest. Our biases make it difficult to discern reliable information when polarization is dominant. And social media platforms have made it easy for ideas—even false ones—to spread around the globe almost instantaneously.

Data scientists have recently made significant progress in understanding the spread and consumption of information, its effect on opinion formation, and the ways people influence one another. Advances in technology have made it possible to exploit the deluge of data from social media—the traces that people use as they choose, share and comment online—to study social dynamics at a high level of resolution.

By applying the methods of computational social science to the traces that people leave on Facebook, Twitter, YouTube and other such outlets, scientists can study the spread of conspiracy theories in great detail. Thanks to these studies, we know that humans are not, as has long been assumed, rational. Presented with unfiltered information, people will appropriate that which conforms to their own thinking. This effect, known as confirmation bias, fuels polarization and thus decreases the value of information that is true—propping up theories about global mega conspiracies, connections between vaccines and autism, and other nonsense. On top of this, as the COVID-19 pandemic has made clear, the nuances of how science is conducted are not well understood by the public, and concepts such as uncertainty and complexity are difficult to convey. And some scientists are not very good at communicating.

THE ECHO CHAMBER

AT THE UNIVERSITY OF VENICE, my colleagues and I have spent the past year investigating the spread of information and misinformation on social networks. We are especially interested in learning how information goes viral and how opinions are formed and reinforced in cyberspace.

One of our first studies on the subject was designed to reveal how social media users treat three different types of information: mainstream news, alternative news and online political activism. The first category is self-explanatory: it refers to the media outlets that provide nationwide news coverage in Italy. The second category includes outlets that claim to report information that the mainstream media have "hidden." The final category refers to content published by activist groups that use the Web as a tool for political mobilization.

Gathering information for our study, especially from alternative sources, was time-consuming and painstaking. We collected and manually verified various indicators from Facebook users and groups active in fact-checking. From the 50 Facebook pages that we investigated, we analyzed the online behavior of more than two million Italian users who interacted with those pages between September 2012 and February 2013. We found that posts on qualitatively different topics behaved very similarly online: the same number of users tended to interact with them, to share them on social media and to debate them. In other words, information from major daily newspapers, alternative news sources and political activist sites all reverberate in the same way.

Two different hypotheses could explain this result. The first possibility was that *all* users treat *all* information equally, regardless of its veracity. The other was that members of certain interest groups treat all information equally, whether it is true or not, *if it reinforces their preexisting beliefs*. The second hypothesis was, to us, more interesting. It suggested that confirmation bias plays an important role in the spread of misinformation.

THE MANAGER AND THE MESSAGE

THE NEXT STEP was to test these two hypotheses. We decided to compare the online behavior of people who read science news with that of people who usually follow alternative news and conspiracy theories. We chose these two types of content because of a very specific difference—whether they have a sender, a manager of the message. Science news is about studies published in

© 2020 Scientific American

scientific journals—work by authors and institutions that are known quantities. Conspiracy theories, in contrast, have no sender: they are formulated in a way to promote the uncertainty they generate. The subject is always a secret plan or truth that someone is deliberately concealing from the public.

The very complexity of our world—issues such as multiculturalism, the growing intricacy of the global financial system and technological progress—can lead people, regardless of educational level, to choose to believe compact explanations that clearly identify an object of blame. Martin Bauer, a social psychologist at the London School of Economics and a scholar of conspiracy dynamics, describes conspiracy thinking as a "quasireligious mentality." It is a little bit like the dawn of humanity, when people attributed storms to acts of divinity.

For this study, which we called "Science vs Conspiracy: Collective Narratives in the Age of Misinformation" and published in *PLOS ONE,* we investigated 73 Facebook pages, 39 of which trafficked in conspiracy and 34 of which published science news. Altogether, these pages had more than a million Italian users between 2010 and 2014. We found that both sets of pages attracted very attentive audiences—users who rarely leave their echo chambers. People who read science news rarely read conspiracy news, and vice versa. But the conspiracy pages attracted three times more users.

Facebook's news feed algorithm seems to foment the formation of echo chambers, which play an important role in the spread of false rumors. When we investigated 4,709 posts that satirized conspiracy theories (example: "Airplane chemtrails contain Viagra"), we found that consumers of "real" conspiracy news were much more likely to read these satirical pieces than readers of legitimate science news. We also found that users who focus primarily on conspiracy news tend to share content more widely.

When we reconstructed the social networks of our two groups (science news readers and fans of conspiracy theories), we discovered a surprising statistical regularity: as the number of likes for a specific type of narrative increased, the probability of having a virtual social network composed *solely* of users with the same profile also increased. In other words, the more you are exposed to a certain type of narrative, the greater the probability that *all* your Facebook friends will have the same news preferences. The division of social networks into homogeneous groups is crucial to understanding the viral nature of the phenomenon. These groups tend to exclude anything that does not fit with their worldview. Such a tendency has been confirmed in more general settings: how people read newspapers on Facebook. Our study, published online in February 2017 in the *Proceedings of the National Academy of Sciences USA,* analyzing the behavior of 376 million users interacting with news outlet pages on social media, found that the more a user is active online the fewer number of sources the person engages with.

A WICKED PROBLEM

IN 2014 WE DECIDED to start investigating efforts to correct the spread of unsubstantiated claims on social media. Does debunking work? To find out, we measured the "persistence"—the tendency of a person to continue engaging with a specific type of content over time—of conspiracy news readers who had been exposed to debunking campaigns. People who were exposed to debunking campaigns were 30 percent more likely to keep reading conspiracy news. In other words, for a certain type of user, debunking actually reinforces belief in the conspiracy.

We observed the same dynamics in a study of 55 million Facebook users in the U.S. Users avoid cognitive discord by consuming information that supports their preexisting beliefs, and they share that information widely. Moreover, we found that over time people who embrace conspiracy theories in one domain—say, the (nonexistent) connection between vaccines and autism—will seek out such theories in other domains. Once inside the echo chamber, they tend to embrace the entire conspiracy corpus.

These dynamics suggest that the spread of online misinformation will be very hard to stop. Any attempt at reasoned discussion usually degenerates into a fight between extremists, which ends in polarization. In this context, it is quite difficult to accurately inform people and almost impossible to stop a baseless report. We performed another experiment to measure the effect of different journalistic techniques in smoothing polarization, but those tactics proved to be relatively ineffective.

THE COMPLEXITY OF OUR WORLD CAN LEAD PEOPLE TO CHOOSE TO BELIEVE COMPACT EXPLANATIONS THAT CLEARLY IDENTIFY AN OBJECT OF BLAME.

In all probability, social media will continue to teem with debates on the latest global mega conspiracy. The important thing is to share what is being hidden from us; whether it is true or false hardly matters. Perhaps we should stop calling this the Information Age and start calling it the Age of Segregation. SA

Walter Quattrociocchi coordinates the Laboratory of Data Science and Complexity at the University of Venice. His research focuses on the quantitative characteristics of social dynamics, from how opinions are formed to the way information spreads. Quattrociocchi is particularly interested in the origin, production, and spread of online narratives and social contagion.

MORE TO EXPLORE

Anatomy of News Consumption on Facebook. Ana Lucía Schmidt et al. in *Proceedings of the National Academy of Sciences USA,* Vol. 114, No. 12, pages 3035–3039; March 21, 2017.
Debunking in a World of Tribes. Fabiana Zollo et al. in *PLOS ONE,* Article 0181821; July 24, 2017.
Social and Political Challenges. Part 2 : The Global Risks Report 2017. 12th edition. World Economic Forum, 2017. Available at http://reports.weforum.org/global-risks-2017
Echo Chambers on Social Media: A Comparative Analysis. Matteo Cinelli et al. Submitted to arXiv.org April 20, 2020. Preprint available at arxiv.org/abs/2004.09603
Measuring Social Response to Different Journalistic Techniques on Facebook. Ana L. Schmidt et al. in *Humanities and Social Sciences Communications,* Vol. 7, No. 17, 7 pages; July 1, 2020.

scientificamerican.com/magazine/sa

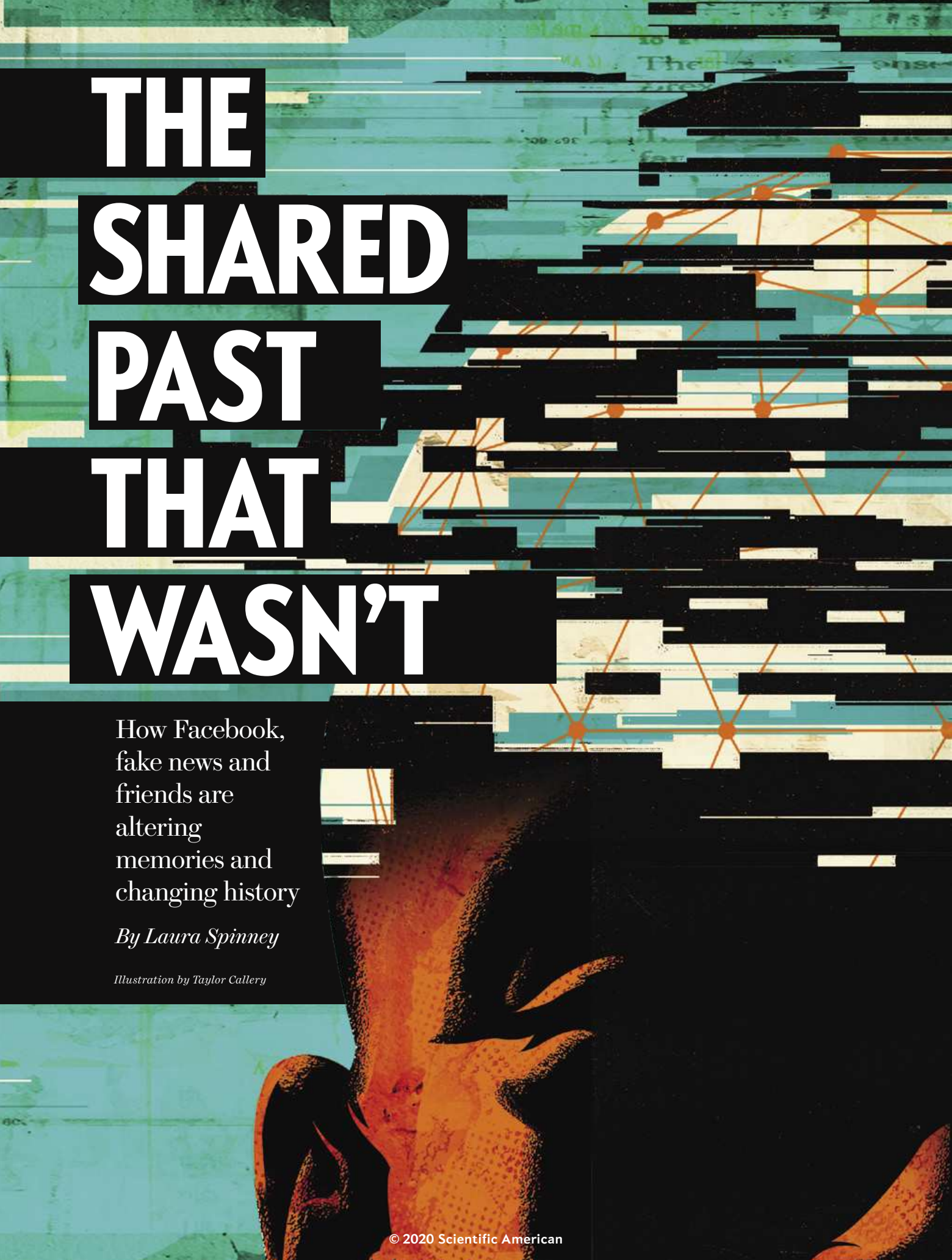
© 2020 Scientific American

THE SHARED PAST THAT WASN'T

How Facebook, fake news and friends are altering memories and changing history

By Laura Spinney

Illustration by Taylor Callery



© 2020 Scientific American

© 2020 Scientific American

STRANGE THINGS HAVE BEEN HAPPENING IN THE NEWS lately. In 2017 members of President Donald Trump's administration alluded to a "Bowling Green massacre" and terror attacks in Sweden and Atlanta that never happened. The misinformation was swiftly corrected, but some historical myths have proved difficult to erase. Since at least 2010, for example, an online community has shared the apparently unshakable recollection of Nelson Mandela dying in prison in the 1980s, despite the fact that he lived until 2013, leaving prison in 1990 and going on to serve as South Africa's first Black president.

Memory is notoriously fallible, but some experts worry that a new phenomenon is emerging. "Memories are shared among groups in novel ways through sites such as Facebook and Instagram, blurring the line between individual and collective memories," says psychologist Daniel Schacter, who studies memory at Harvard University. "The development of Internet-based misinformation, such as recently well-publicized fake news sites, has the potential to distort individual and collective memories in disturbing ways."

Collective memories form the basis of history, and people's understanding of history shapes how they think about the future. The fictitious terrorist attacks, for example, were cited to justify a travel ban on the citizens of seven "countries of concern." Although history has frequently been interpreted for political ends, psychologists are now investigating the fundamental processes by which collective memories form, to understand what makes them vulnerable to distortion. They show that social networks powerfully shape memory and that people need little prompting to conform to a majority recollection—even if it is wrong. Not all the findings are gloomy, however. Research is pointing to ways of dislodging false memories or preventing them from forming in the first place.

To combat the influence of fake news, says Micah Edelson, a memory researcher at the University of Zurich in Switzerland, "it's important to understand not only the creation of these sites but also how people respond to them."

ALL TOGETHER NOW

COMMUNICATION SHAPES MEMORY. Research on pairs of people conversing about the past shows that a speaker can reinforce aspects of an event by selectively repeating them. That makes sense. Things that get mentioned get remembered—by both speaker and listener. There is a less obvious corollary: related information that goes unmentioned is more likely to fade than unrelated material, an effect known as retrieval-induced forgetting.

These cognitive, individual-level phenomena have been proposed as a mechanism for memory convergence—the process by which two or more people come

© 2020 Scientific American

to agree on what happened. But in the past few years clues have emerged that group-level forces influence convergence, too. In 2015 psychologists Alin Coman of Princeton University and William Hirst of the New School for Social Research reported that a person experiences more induced forgetting when listening to someone in their own social group—a student at the same university, for example—than if they see that person as an outsider. That is, memory convergence is more likely to occur within social groups than between them—an important finding in light of survey data suggesting that 62 percent of U.S. adults get their news from social media, where group membership is often obvious and reinforced..

Groups can also distort memories. In 2011 Edelson, then at the Weizmann Institute of Science in Rehovot, Israel, showed 30 volunteers a documentary. They watched the film in groups of five and, a few days later, answered questions about it individually. One week after the viewing session, participants answered questions again—but only after seeing answers that members of their group had supposedly given. When most of the fabricated responses were false, participants conformed to the same false answer about 70 percent of the time—despite having initially responded correctly. But when they learned that the answers had been generated randomly, the participants reversed their incorrect answers only about 60 percent of the time. "We found that processes that happen during initial exposure to erroneous information make it more difficult to correct such influences later," Edelson says.

Studying those processes as they happen—as collective memories are shaped through conversation—has been difficult to do in large groups. Years ago monitoring communication in groups of 10 or more would have required several rooms for private conversations, many research assistants and lots of time. Now multiple participants can interact digitally in real time. Coman's group has developed a software platform that can track exchanges between volunteers in a series of timed chats. "It takes one research assistant 20 minutes and one lab room," Coman says.

In 2016 the group used this software to ask, for the first time, how the structure of social networks affects the formation of collective memories in large groups. The researchers fed information about four fictional Peace Corps volunteers to 140 participants from Princeton University, divided into groups of 10. First, the participants were asked to recall as much information as they could on their own. Then they took part in a series of three conversations—online chat sessions lasting a few minutes each—with other members of their group, in which they recalled the information collaboratively. Finally, they tried to recall the events individually again.

The researchers investigated two scenarios—one in which the group formed two subclusters, with almost all conversations taking place within the subclusters, and one in which it formed one large cluster [*see box on next page*]. Although people in the single cluster agreed on the same set of information, Coman says, those in the two subclusters generally converged on different "facts" about the fictional volunteers.

This effect is evident in real-world situations. Palestinians living in Israel and those in the West Bank, who were separated by force during the Arab-Israeli wars of 1948 and 1967, have gravitated to different versions of their past, despite a shared Arab-Palestinian identity. Similarly divergent truths emerged after the erection of the Berlin Wall.

In the lab, Coman can manipulate social networks and look at the memories that form. His comparison of the two scenarios revealed the importance of "weak links" in information propagation. These are links between, rather than within, networks—acquaintances, say, rather than friends—and they help to synchronize the versions held by separate networks. "They are probably what drives the formation of community-wide collective memories," he says.

"MEMORIES ARE SHARED AMONG GROUPS IN NOVEL WAYS THROUGH SITES SUCH AS FACEBOOK AND INSTAGRAM, BLURRING THE LINE BETWEEN INDIVIDUAL AND COLLECTIVE MEMORIES."

—DANIEL SCHACTER, HARVARD UNIVERSITY

One function of those weak links might be to remind people of information expunged through the processes of memory convergence. But timing is important. Coman has shown that information introduced by a weak link is much more likely to shape the network's memory if it is introduced before its members talk among themselves. Once a network agrees on what happened, collective memory becomes relatively resistant to competing information.

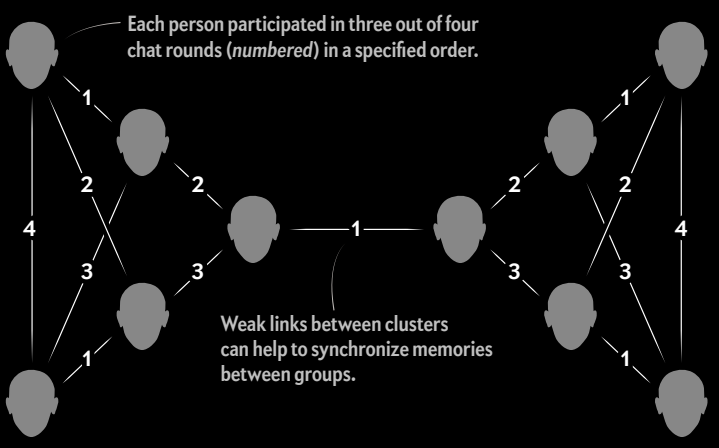
Coman thinks that memory convergence bolsters group cohesion. "Now that we share a memory, we can have a stronger identity and might care more about each other," he says. Abundant research links strong group identity with higher reported individual well-being. This is shown by research on the family. At Emory University, psychologist Robyn Fivush is studying the stories that families tell themselves. "What we find is that adolescents and young adults who know more family stories show better psychological well-being," she says.

Although shared memories may foster more closely knit groups, they can also distort the role of outsiders, driving a wedge between groups. Memory shapes

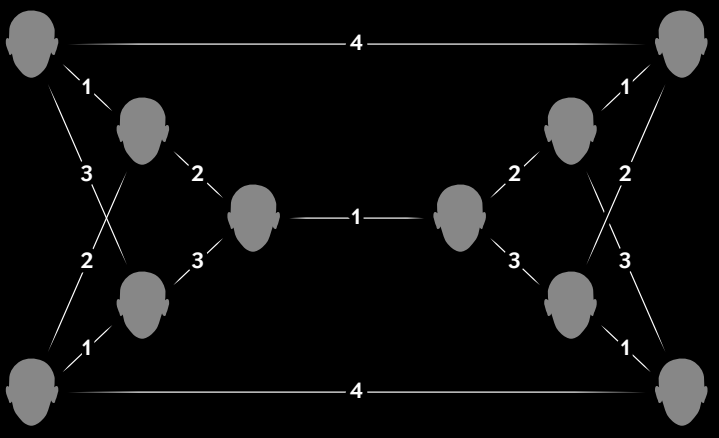
© 2020 Scientific American

Hello, Operator

In an experiment, 10 **volunteers** discussed the details of a story in one-on-one chats with three other people. The distance between participants in the network (degrees of separation) correlated to how well their recollection of the story aligned with one another.


Each person participated in three out of four chat rounds (*numbered*) in a specified order.
Weak links between clusters can help to synchronize memories between groups.

CLUSTERED
With as many as five degrees of separation between some participants, memories converged within clusters but less so between them.



UNCLUSTERED
With no more than three degrees of separation, there was more alignment in the memories of any two members.

group identity, which in turn shapes memory, in a potentially vicious cycle. Weak links have an important corrective effect, but in their absence, two groups might converge on mutually incompatible versions of the past. These may be preserved for posterity in statues and history books. But they can evolve over time.

MAKING MEMORIES, MAKING HISTORIES

IN OSTEND, BELGIUM, a public monument depicts Belgiam king Leopold II, surrounded by two groups of grateful subjects—one Belgian, the other Congolese. In 2004 protesters, who felt that the monument misrepresented history, severed the bronze hand of one of the Congolese figures. They explained anonymously to a local newspaper that the amputation more accurately reflected Leopold's role in Belgium's African colony: not genial protector but brutal tyrant.

In 2010 social psychologists Laurent Licata and Olivier Klein of the Free University of Brussels carried out a survey to explore different generations' attitudes toward Belgium's colonial past. They found that Belgian students expressed higher levels of collective guilt and support for reparative actions toward what is now the Democratic Republic of the Congo than did their parents, who in turn expressed higher levels than their parents. An important factor shaping that evolution, the researchers suggest, was Adam Hochschild's influential book *King Leopold's Ghost* (Houghton Mifflin, 1998), which painted a much darker picture of the colonial period than had been accepted previously. "Those who were young when that book came out were particularly marked by it," Licata says, "whereas older Belgians had grown up with a different set of facts."

Not all collective memories pass into history. Cognitive psychologists Norman Brown of the University of Alberta and Connie Svob of Columbia University have proposed that something besides cognitive and social processes determines whether an event survives the transition across generations: the nature of the event itself. "It is the amount of change to a person's fabric of daily life that is most crucially at stake," Svob says.

In a study published in 2016, they reported that the children of Croatians who had lived through the Yugoslav wars of the 1990s were more likely to recall their parents' war-related experiences—getting shot, for example, or the house being bombed—than their non-war-related ones, such as marriage or birth of a first child. Wars, as with immigration, bring great upheaval in their wake and so are highly memorable, Svob says.

This "transition theory" she says, could also explain one of the biggest voids in Westerners' collective memory of the 20th century—why they easily recall the two World Wars but not the flu pandemic of 1918–1920 that killed more than either of them (considering unrecorded deaths from the disease). "The degree of change wrought by war tends to be greater than the degree of change wrought by a pandemic," Svob says. Others find that explanation puzzling: "If you lost a loved one in the flu epidemic," Fivush says, "then it certainly disrupted your daily life."

The set of collective memories that a group holds

NATURE; SOURCE: "MNEMONIC CONVERGENCE IN SOCIAL NETWORKS: THE EMERGENT PROPERTIES OF COGNITION AT A COLLECTIVE LEVEL," BY A. COMAN, I. MOMENNEJAD, R. D. DRACH AND A. GEANA, IN *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES USA*, VOL. 113, NO. 29; JULY 19, 2016

© 2020 Scientific American

clearly evolves over time. One reason for this is that people tend to be marked most by events in their adolescence or young adulthood—a phenomenon known as the reminiscence bump. As a new generation grows up, events that happen to its members during their youth override the events that previously dominated society and thus "update" the collective memory. A 2016 survey by the Pew Research Center in Washington, D.C., showed that the defining historical moments for baby boomers in the U.S. were the terror attacks on September 11, 2001, and the assassination of John F. Kennedy. For those born since 1965, they were the attacks on 9/11 and the 2008 election of President Barack Obama.

And over time each generation adds some events and forgets others. Psychologists Henry Roediger of Washington University in St. Louis and Andrew DeSoto of the Association for Psychological Science in Washington, D.C., report, for example, that successive U.S. generations forget their past presidents in a regular manner that can be described by a power function. They predict that Harry Truman (1945–1953) will be as forgotten by 2040 as William McKinley (1897–1901) is today.

That evolution is reflected by evolving attitudes toward the future. Roediger and anthropologist James Wertsch, also at Washington University, have observed that U.S. politicians debating the invasion of Iraq in the early 2000s fell into two groups: those who advocated invasion on the grounds that Saddam Hussein had to be stopped like Adolf Hitler before him and those who opposed it because they feared another bloody, protracted Vietnam War. Although each might have chosen their historical precedent for political reasons, they in turn reinforced that precedent in the memory of anyone who heard them speak.

SPOTTING THE FAKE

RESEARCH INTO COLLECTIVE MEMORY is pointing to ways that it might be shaped for the collective good. Edelson and his team gave grounds for optimism when, in a 2014 follow-up to their earlier study, they reported that although some false memories are resistant to change, the people who hold them can nonetheless be influenced by credible information. The team used functional magnetic resonance imaging to scan volunteers' brains as they recalled information about a film. The scans revealed changes in brain activation that correlated with the degree of confidence in an inaccurate memory—and, ultimately, with whether they reverted to their initial, accurate one. "By exposing them to the fact that this information is not credible, in most cases, individuals will take that into account," Edelson says. "In 60 percent of cases, they will flip their answer. But even if they maintain a wrong answer, they'll be less confident about it."

Coman has two suggestions from his findings. The first is directed at the justice system. In some U.S. states, jurors are forbidden to take notes made during a trial into the deliberation room—a legacy of historically high illiteracy rates and a belief that the group remembers more reliably than the individual. In fact, Coman says, using notes could protect jurors from retrieval-induced biases and group-level social influences.

"IF YOU UNDERSTAND THE NATURE OF THE FALSE INFORMATION, YOU CAN TARGET IT FOR SUPPRESSION."
—ALIN COMAN, PRINCETON UNIVERSITY

His second suggestion concerns the diffusion of crucial information to the public during emergencies such as epidemics. Having observed that retrieval-induced forgetting is enhanced in high-anxiety situations, Coman has come up with some advice for officials: draw up a short but comprehensive list of key points, make sure that all officials have the same list, repeat those points often and keep tabs on bad information that enters circulation. During the 2014 Ebola outbreak, for example, concerns in the U.S. were stoked by a misconception that being in the same room as a person with the infection was enough to catch it. The best way to kill that rumor, Coman says, would have been to explain—often—that Ebola could be transmitted only through bodily fluids. "If you understand the nature of the false information, you can target it for suppression just by mentioning information that is conceptually related but accurate," he says.

Collective memory is a double-edged sword. Some will no doubt use it to mislead. "The fact that information can freely circulate in the community has been considered one of the most important and constructive features of open and democratic societies," Coman says. "But creating such societies does not inherently guarantee positive outcomes." False collective memories might be the price of defending free speech. But understanding how they form might offer some protection the next time people are reminded about a massacre that never happened. SA

Laura Spinney is a science journalist based in Paris.

MORE TO EXPLORE

Intergenerational Transmission of Historical Memories and Social-Distance Attitudes in Post-War Second-Generation Croatians. Connie Svob et al. in *Memory and Cognition*, Vol. 44, pages 846–855; March 29, 2016.

Americans Name the 10 Most Significant Historic Events of Their Lifetimes. Claudia Deane, Maeve Duggan and Rich Morin. December 15, 2016. Available at www.journalism.org

scientificamerican.com/magazine/sa

© 2020 Scientific American

© 2020 Scientific American

CLICKS, LIES AND VIDEOTAPE

Artificial intelligence is making it possible for anyone to manipulate audio and video. The biggest threat is that we stop trusting anything at all

By Brooke Borel

IN BRIEF

Rapidly evolving AI technologies allow for the automated creation of fake video and audio. Some experts worry that the spread of disinformation via social media could have profound effects on public discourse and political stability.

Computer scientists are working on AI detection tools to flag fake videos, but they lag behind the ability to create manipulated content. Meanwhile social scientists warn that policing fakes post hoc is not a sufficient solution.

Written fake news was a troubling factor in the 2016 U.S. elections. Research suggests that fake video may be especially effective at stoking fear—an emotion that powers viral content. One concern is that it could erode our trust in all media, including what is real.

Illustration by Taylor Callery

© 2020 Scientific American

IN APRIL 2018 A NEW VIDEO OF BARACK OBAMA SURFACED ON THE INTERNET. Against a backdrop that included both the American and presidential flags, it looked like many of his previous speeches. Wearing a crisp white shirt and dark suit, Obama faced the camera and punctuated his words with outstretched hands: "President Trump is a total and complete dipshit."

Without cracking a smile, he continued. "Now, you see, I would never say these things. At least not in a public address. But someone else would." The view shifted to a split screen, revealing the actor Jordan Peele. Obama hadn't said anything—it was a real recording of an Obama address blended with Peele's impersonation. Side by side, the message continued as Peele, like a digital ventriloquist, put more words in the former president's mouth.

In this era of fake news, the video was a public service announcement produced by BuzzFeed News, showcasing an application of new artificial-intelligence (AI) technology that could do for audio and video what Photoshop has done for digital images: allow for the manipulation of reality.

The results are still fairly unsophisticated. Listen and watch closely, and Obama's voice is a bit nasally. For brief flashes, his mouth—fused with Peele's—floats off-center. But this rapidly evolving technology, which is intended for Hollywood film editors and video game makers, has the imaginations of some national security experts and media scholars running dark. The next generation of these tools may make it possible to create convincing fakes from scratch—not by warping existing footage, as in the Obama address, but by orchestrating scenarios that never happened at all.

The consequences for public knowledge and discourse could be profound. Imagine, for instance, the impact on the upcoming elections if a fake video smeared a politician during a tight race. Or attacked a CEO the night before a public offering. A group could stage a terrorist attack and fool news outlets into covering it, sparking knee-jerk retribution. Even if a viral video is later proved to be fake, will the public still believe it was true anyway? And perhaps most troubling: What if the very idea of pervasive fakes makes us stop believing much of what we see and hear—including the stuff that is real?

Many technologists acknowledge the potential for sweeping misuse of this technology. But while they fixate on "sexy solutions for detection and disclosure, they spend very little time figuring out whether any of that actually has an effect on people's beliefs on the validity of fake video," says Nate Persily, a law professor at Stanford University. Persily studies, among other topics, how the Internet affects democracy, and he is among a growing group of researchers who argue that curbing viral disinformation cannot be done through technical fixes alone. It will require input from psychologists, social scientists and media experts to help tease out how the technology will land in the real world.

"We've got to do this now," Persily says, "because at the moment the technologists—necessarily—drive the discussion" on what may be possible with AI-generated video. Already our trust in democratic institutions such as government and journalism is ebbing. With social media a dominant distribution channel for information, it is even easier today for fake-news makers to exploit us. And with no cohesive strategy in place to confront an increasingly sophisticated technology, our fragile collective trust is even more at risk.

INNOCUOUS BEGINNINGS

THE PATH TO FAKE VIDEO traces back to the 1960s, when computer-generated imagery was first conceived. In the 1980s these special effects went mainstream, and ever since, movie lovers have watched the technology evolve from science-fiction flicks to Forrest Gump shaking hands with John F. Kennedy in 1994 to the revival of Peter Cushing and Carrie Fisher in *Rogue One*. The goal has always been to "create a digital world where any storytelling could be possible," says Hao Li, an associate professor of computer science at the University of Southern California and CEO of Pinscreen, an augmented-reality start-up. "How can we create something that appears real, but everything is actually virtual?"

Early on, most graphics came from artists, who used computers to create three-dimensional models and then hand-painted textures and other details—a tedious process that did not scale up. About 20 years ago some computer-vision researchers started thinking of graphics differently: rather than spending time on individual models, why not teach computers to create from data? In 1997 scientists at the Interval Research Corporation in Palo Alto, Calif., developed Video Rewrite, which sliced up existing footage and reconfigured it. The researchers made a clip of JFK saying, "I never met Forrest Gump." Soon after, scientists at the Max Planck Institute for Biological Cybernetics in Tübingen, Germany, taught a computer to pull features from a data set of 200 three-dimensional scans of human faces to make a new face.

The biggest recent jump in the relationship among computer vision, data and automation arguably came in 2012, with advances

© 2020 Scientific American

TECHNOLOGY that was originally developed to create virtual scenes in film (*1*) has evolved into a tool that can be used to make fake videos (*2*) to spread disinformation.

FILM STILL FROM *FORREST GUMP*. PARAMOUNT PICTURES, 1994 (*1*); FILM STILL FROM *YOU WON'T BELIEVE WHAT OBAMA SAYS IN THIS VIDEO!* MONKEYPAW PRODUCTIONS AND BUZZFEED, APRIL 17, 2018 (*2*)

in a type of AI called deep learning. Unlike the work from the late 1990s, which used static data and never improved, deep learning adapts and gets better. This technique reduces objects, such as a face, to bits of data, says Xiaochang Li, now in the department of communication at Stanford. "This is the moment where engineers say: we are no longer going to model things," she says. "We are going to model our ignorance of things and just run the data to understand patterns."

Deep learning uses layers of simple mathematical formulas called neural networks, which get better at a task over time. For example, computer scientists can teach a deep-learning tool to recognize human faces by feeding it hundreds or thousands of photographs and essentially saying, each time, *this is a face* or *this is not a face*. Eventually when the tool encounters a new person, it will recognize patterns that make up human features and say, statistically speaking, *this is also a face*.

Next came the ability to concoct faces that looked like real people, using deep-learning tools known as generative networks. The same logic applies: computer scientists train the networks on hundreds or thousands of images. But this time the network follows the patterns it gleaned from the examples to make a new face. Some companies are now using the same approach with audio. In 2018 Google unveiled Duplex, an AI assistant based on software called WaveNet, which can make phone calls and sounds like a real person—complete with verbal tics such as uhs and hmms. In the future, a fake video of a politician may not need to rely on impersonations from actors like Peele. In April 2017 Lyrebird, a Canadian start-up, released sample audio that sounded creepily like Obama, Trump and Hillary Clinton.

But generative networks need big data sets for training, and that can require significant human labor. The next step in improving virtual content was to teach the AI to train itself. In 2014 researchers at the University of Montreal did this with a generative adversarial network, or GAN, which puts two neural networks in conversation. The first is a generator, which makes fake images, and the second is a discriminator, which learns to distinguish between real and fake. With little to no human supervision, the networks train one another through competition—the discriminator nudges the generator to make increasingly realistic fakes, while the generator keeps trying to trick the discriminator. GANs can craft all sorts of stuff. At the University of California, Berkeley, scientists built one that can turn images of horses into zebras or transform Impressionist paintings by the likes of Monet into crisp, photorealistic scenes.

Then, in May 2018, researchers at the Max Planck Institute for Informatics in Saarbrücken, Germany, and their colleagues revealed "deep video," which uses a type of GAN. It allows an actor to control the mouth, eyes and facial movements of someone else in prerecorded footage. Deep video currently only works in a portrait setup, where a person looks directly at the camera. If the actor moves too much, the resulting video has noticeable digital artifacts such as blurred pixels around the face.

GANs are not yet capable of building complex scenes in video that are indistinguishable from ones captured in real footage. Sometimes GANs produce oddities, such as a person with an eyeball growing out of his or her forehead. In February 2018, however, researchers at the company NVIDIA figured out a way to get GANs to make incredibly high-resolution faces by starting the training on relatively small photographs and then building up the resolution step by step. And Hao Li's team at the University of Southern California has used GANs to make realistic skin, teeth and mouths, all of which are notoriously difficult to digitally reconstruct.

© 2020 Scientific American

None of these technologies are easy for nonexperts to use well. But BuzzFeed's experiment hints at our possible future. The video came from free software called FakeApp—which used deep learning, though not GAN. The resulting videos are dubbed deepfakes, a mash-up of "deep learning" and "fake," named after a user on the Web site Reddit, who, along with others, was an early adopter and used the tech to swap celebrities' faces into porn. Since then, amateurs across the Web have used FakeApp to make countless videos—most of them relatively harmless pranks, such as adding actor Nicolas Cage to a bunch of movies he was not in or morphing Trump's face onto the body of German chancellor Angela Merkel. More ominous are the implications. Now that the technology is democratized, anyone with a computer can hypothetically use it.

CONDITIONS FOR FAKE NEWS

EXPERTS HAVE LONG WORRIED that computer-enabled editing would ruin reality. Back in 2000, an article in *MIT Technology Review* about products such as Video Rewrite warned that "seeing is no longer believing" and that an image "on the evening news could well be a fake—a fabrication of fast new video-manipulation technology." Twenty years later fake videos don't seem to be flooding news shows. For one thing, it is still hard to produce a really good one. It took 56 hours for BuzzFeed to make the Obama clip with help from a professional video editor.

The way we consume information, however, has changed. Today only about half of American adults watch the news on television, whereas two thirds get at least some news via social media, according to the Pew Research Center. The Internet has allowed for a proliferation of media outlets that cater to niche audiences—including hyperpartisan Web sites that intentionally stoke anger, unimpeded by traditional journalistic standards. The Internet rewards viral content that we are able to share faster than ever before, Persily says. And the glitches in fake video are less discernible on a tiny mobile screen than a living-room TV.

"We will not win this game. It's just that we will make it harder and harder for the bad guys to play it."
—Alexei Efros *University of California, Berkeley*

The question now is what will happen if a deepfake with significant social or political implications goes viral. With such a new, barely studied frontier, the short answer is that we do not know, says Julie Carpenter, a research fellow with the Ethics + Emerging Sciences Group, based at California State Polytechnic University, San Luis Obispo, who studies human-robot interaction. It is possible we will find out soon enough, with key elections coming up this fall in the U.S., as well as internationally.

We have already witnessed the fallout when connectivity and disinformation collide. Fake news—fabricated text stories designed to look like legitimate news reports and to go viral—was a much discussed feature of the 2016 U.S. presidential election. According to collaborative research from Princeton University, Dartmouth College and the University of Exeter in England, roughly one in four Americans visited a fake news site during the five weeks between October 7 and November 14, 2016, mostly through the conduit of their Facebook feeds. Moreover, 2016 marked a low point in the public's trust in journalism. By one estimate, just 51 percent of Democrats and 14 percent of Republicans said they trusted mass media.

The science on written fake news is limited. But some research suggests that seeing false information just once is sufficient to make it seem plausible later on, says Gordon Pennycook, an assistant professor of organizational behavior at the University of Regina in Saskatchewan. It is not clear why, but it may be thanks to "fluency," he says, or "the ease at which it is processed." If we hear Obama call Trump a curse word and then later encounter another false instance where Obama calls Trump obscene names, we may be primed to think it is real because it is familiar.

According to a study from the Massachusetts Institute of Technology that tracked 126,000 stories on Twitter between 2006 and 2017, we are also more likely to share fake news than real news—and especially fake political stories, which spread further and quicker than those about money, natural disasters or terrorism. The paper suggested that people crave novelty. Fake news in general plays to our emotions and personal identity, enticing us to react before we have had a chance to process the information and decide if it is worth spreading. The more that content surprises, scares or enrages us, the more we seem to share it.

There are troubling clues that video may be especially effective at stoking fear. "When you process information visually, you believe that this thing is closer to you in terms of space, time or social group," says Elinor Amit, an assistant professor of cognitive, linguistic and psychological sciences at Brown University, whose work teases out the differences in how we relate to text and images. She hypothesizes that this distinction is evolutionary—our visual development came before written language, and we rely more on our senses to detect immediate danger.

Fake video has, in fact, already struck political campaigns. In July 2018 Allie Beth Stuckey, a TV host at Conservative Review, posted on Facebook an interview with Alexandria Ocasio-Cortez, a Democratic congressional nominee from New York City. The video was not a deepfake but an old-fashioned splice of a real interview with new questions to make Ocasio-Cortez appear to flub her answers. Depending on your political persuasion, the video was either a smear job or, as Stuckey later called it in her defense, satire. Either way, it had 3.4 million views within a week and more than 5,000 comments. Some viewers seemed to think Ocasio-Cortez had bombed a real interview. "Omg! She doesn't know what and how to answer," one wrote. "She is stupid."

That all of this is worrying is part of the problem. Our dark ruminations may actually be worse for society than the videos themselves. Politicians could sow doubt when their real misdeeds are caught on tape by claiming they were faked, for example. Knowing that convincing fakes are even possible might erode our trust in all media, says Raymond J. Pingree, an associate professor in mass communications at Louisiana State University. Pingree studies how confident people are in their ability to evaluate

© 2020 Scientific American

what is real and what is not and how that affects their willingness to participate in the political process. When individuals lose that confidence, they are more likely to fall for liars and crooks, he says, and "it can make people stop wanting to seek the truth."

A GAME OF CAT AND MOUSE

TO A COMPUTER SCIENTIST, the solution to a bug is often just more computer science. Although the bugs in question here are far more complex than bad coding, there is a sense in the community that algorithms could be built to flag the fakes.

"There is certainly technical progress that can be made against the problem," says R. David Edelman of M.I.T.'s Internet Policy Research Initiative. Edelman, who served as a tech adviser under Obama, has been impressed by faked videos of the former president. "I know the guy. I wrote speeches for him. I couldn't tell the difference between the real and fake video," he says. But while he could be fooled, Edelman says, an algorithm might pick up on the "telltale tics and digital signatures" that are invisible to the human eye.

So far the fixes fall within two categories. One proves that a video is real by embedding digital signatures, analogous to the intricate seals, holograms and other features that currency printers use to thwart counterfeiters. Every digital camera would have a unique signature, which, theoretically, would be tough to copy.

The second strategy is to automatically flag fake videos with detectors. Arguably the most significant push for such a detector is a program from the Defense Advanced Research Projects Agency called Media Forensics, or MediFor. It kicked off in 2015, not long after a Russian news channel aired fake satellite images of a Ukrainian fighter jet shooting at Malaysia Airlines Flight 17. Later, a team of international investigators pegged the flight's downing on a Russian missile. The satellite images were not made with deep learning, but DARPA saw the coming revolution and wanted to find a way to fight it, says David Doermann, MediFor's former program manager.

MediFor is taking three broad approaches, which can be automated with deep learning. The first examines a video's digital fingerprint for anomalies. The second ensures a video follows the laws of physics, such as sunlight falling the way it would in the real world. And the third checks for external data, such as the weather on the day it was allegedly filmed. DARPA plans to unify these detectors into a single tool, which will give a point score on the likelihood that a video is fake.

These strategies could cut down on the volume of fakes, but it will still be a game of cat and mouse, with forgers imitating digital watermarks or building deep-learning tools to trick the detectors. "We will not win this game," says Alexei Efros, a professor of computer science and electrical engineering at U.C. Berkeley, who is collaborating with MediFor. "It's just that we will make it harder and harder for the bad guys to play it."

And anyway, these tools are still decades away, says Hany Farid, a professor of computer science at Dartmouth College. As fake video continues to improve, the only existing technical solution is to rely on digital forensics experts like Farid. "There's just literally a handful of people in the world you can talk to about this," he says. "I'm one of them. I don't scale to the Internet."

SAVING REALITY

EVEN IF EACH OF US can ultimately use detectors to parse the Internet, there will always be a lag between lies and truth. That is one reason why halting the spread of fake video is a challenge for the social media industry. "This is as much a distribution problem as it is a creation problem," Edelman says. "If a deepfake falls in the forest, no one hears it unless Twitter and Facebook amplify it."

When it comes to curbing viral disinformation, it is not clear what the legal obligations are for social media companies or whether the industry can be regulated without trampling free speech. Facebook CEO Mark Zuckerberg finally admitted that his platform has played a role in spreading fake news—although it took more than 10 months following the 2016 election. Facebook, after all, was designed to keep users consuming and spreading content, prioritizing what is popular over what is true. With more than two billion active monthly users, it is a tinderbox for anyone who wants to spark an enraging fake story.

Since then, Zuckerberg has promised to act. He is putting some of the burden on users by asking them to rank the trustworthiness of news sources (a move that some see as shirking responsibility) and plans to use AI to flag disinformation. The company has been tight-lipped on the details. Some computer scientists are skeptical about the AI angle, including Farid, who says the promises are "spectacularly naïve." Few independent scientists have been able to study how fake news spreads on Facebook because much of the relevant data has been on lockdown.

Still, all the algorithms and data in the world will not save us from disinformation campaigns if the researchers building fake-video technology do not grapple with how their products will be used and abused after they leave the lab. "This is my plea," Persily says, "that the hard scientists who do this work have to be paired up with the psychologists and the political scientists and the communication specialists—who have been working on these issues for a while." That kind of collaboration has been rare.

In March 2018, however, the Finnish Center for Artificial Intelligence announced a program that will invite psychologists, philosophers, ethicists and others to help AI researchers to grasp the broader social implications of their work. A month later Persily, along with Gary King, a political scientist at Harvard University, launched the Social Data Initiative. The project will, for the first time, allow social scientists to access Facebook data to study the spread of disinformation.

With a responsibility vacuum at the top, the onus of rooting out fake videos is falling on journalists and citizen sleuths. Near the end of the deepfake video of Obama and Peele, both men say: "Moving forward, we need to be more vigilant with what we trust from the Internet. It's a time when we need to rely on trusted news sources." It may have been a fake, but it was true. SA

Brooke Borel is a journalist and author of *The Chicago Guide to Fact-Checking*. She recently competed against an AI fact-checker and won by a worrying margin.

MORE TO EXPLORE

The Science of Fake News. David M. J. Lazer et al. in *Science*, Vol. 359, pages 1094–1096; March 9, 2018.
Why Do People Share Fake News? A Sociotechnical Model of Media Effects. Alice E. Marwick in *Georgetown Law Technology Review*, Vol. 2, No. 2, pages 474–512; 2018.
A Nixon Deepfake, a "Moon Disaster" Speech and an Information Ecosystem at Risk. Video with text by Jeffery DelViscio, July 20, 2020. www.scientificamerican.com/article/a-nixon-deepfake-a-moon-disaster-speech-and-an-information-ecosystem-at-risk1

scientificamerican.com/magazine/sa

© 2020 Scientific American

© 2020 Scientific American

HOW TO DEFRAUD DEMOCRACY

A worst-case cyberwarfare scenario for the 2020 American presidential election

By J. Alex Halderman, as told to Jen Schwartz

J. Alex Halderman is a computer scientist who has shown just how easy it is to hack an election. His research group at the University of Michigan examines how attackers can target weaknesses in voting machinery, infrastructure, polling places and registration rolls, among other features. These days he spends much of his time educating lawmakers, cybersecurity experts and the public on how to better secure their elections. In the U.S., there are still serious vulnerabilities heading into the 2020 presidential contest.

Given the cracks in the system, existing technological capabilities and the motivations of adversaries, Halderman has speculated here on potential cybersecurity disasters that could throw the 2020 election—and democracy itself—into question. Halderman, however, is adamant about one thing: "The only way you can reach certainty that your vote won't be counted is by not casting it. I do not want to scare people off from the polls." What follows is based on two conversations that took place in October 2018 and June 2019; it has been edited and condensed.

THE 2016 U.S. PRESIDENTIAL ELECTION REALLY DID change everything. It caught much of the intelligence and cybersecurity communities off guard and taught us that our threat models for cyberwarfare were wrong. Thanks to the Mueller report, we now know that the Russians made a serious and coordinated effort to undermine the legitimacy of the 2016 election outcome. Their efforts were, I think, far more organized and multipronged than anyone initially realized. And to my knowledge, no state has since done any kind of rigorous forensics on their voting machines to see if they had been compromised. I am quite confident that the Russians will be back in 2020.

I think the intelligence community will continue to try to gain visibility into what malicious actors are planning and what they're doing. It's incredible, really, how much detail has come out of the indictments about specific actions by specific people in the Russian military and leadership. But it's hard to know what we're not seeing. And do we have a parallel level of visibility into North Korea or Iran or China? There are potentially a lot of sophisticated nation-state actors that would want to do us harm in 2020 and beyond.

Since the 2016 elections many states have made improvements to their election machinery, but it's not enough, nor is it happening quickly enough. There are still 40 states that are using voting machines that are at least a decade old, and many of these machines are not receiving software patches for vulnerabilities. Nearly 25 percent of states do not have complete paper trails, so they cannot do postelection auditing of physical ballots. Election security is not a partisan issue. Yet there are roadblocks, especially coming from Republican leadership in the Senate, that make it unlikely that an election security bill is going to advance. I think that is a terrible abdication of Congress's duty to provide for the common defense. Thus, many of the worst-case scenarios for election interference are still going to be possible in 2020.

LEADING UP TO ELECTION DAY

CYBERWARFARE often involves exploiting known vulnerabilities in systems and the basic limits of people's psychology and gullibility. During the primaries and in the months leading up to the election, influence operations on social media are going to get much more precise and data-driven than ever before—and therefore more effective and harder to detect.

Already presidential candidates are finely crafting political advertisements to specific demographics of voters to maximally influence them. So, you might receive one message from a candidate based on what's known about you in consumer databases. And people with slightly different views on certain issues might receive a different message from the same candidate. Of course, the bad guys who are trying to spread outright fictions will begin to harness the same strategy.

As we saw in 2016, one of the goals of attackers is to increase the amount of divisiveness in society—to reduce social cohesion. Suppose the Russians purchase access to the same consumer-pro-

IN BRIEF

There are still major cybersecurity vulnerabilities facing the 2020 U.S. presidential election, in part because the election system is based on faith instead of evidence.

Foreign attackers could target voter-registration rolls and election machinery to either influence the outcome or sow chaos and doubt.

The worst-case scenarios could result in an unprecedented constitutional crisis.

Illustration by Lisk Feng

© 2020 Scientific American

file data that advertisers in political campaigns use to target you. They can combine that with data from political polls and purchased (or stolen) voter-registration lists to figure out exactly how much your individual vote matters and use those tools to push customized disinformation at narrow groups of people. Attackers may even impersonate political candidates. In a crowded Democratic primary season, there will be sweeping opportunity to deploy microtargeted messaging to turn people against one another, even when they agree about most things.

We all assume that more transparency is a good thing. But people have always taken facts out of context when it is helpful to them and harmful to their opponents. Candidates increasingly live with the threat of targeted theft of true information. When information is selectively stolen from particular groups that an attacker wants to disadvantage, the truth can be used as a powerful and one-sided political weapon—and as we saw with the 2016 Hillary Clinton campaign, it was incredibly effective. It is such a fundamental threat to our notions of how the truth in journalism should play out in a democratic process that I'm sure it's going to happen again. And it can get a lot worse than the theft of e-mails. Imagine someone hacking into candidates' smartphones and secretly recording them during private moments or while talking to their aides. My research group is polling political campaigns to assess how well they are protecting themselves from this, and so far I don't think they are ready.

SINCE THE 2016 ELECTIONS MANY STATES HAVE MADE IMPROVEMENTS TO THEIR ELECTION MACHINERY, BUT IT'S NOT ENOUGH, NOR IS IT HAPPENING QUICKLY ENOUGH.

We're also going to see information that is doctored or entirely synthetic and made to appear real. In some ways, this creates a worse threat. Attackers don't have to actually catch the candidate saying something or e-mailing something if they can produce a record that is indistinguishable from the truth. We've seen recent advances in using machine learning to synthesize video of people saying things that they never actually said on camera. Overall, these tactics help to undermine our basic notions of what's true and what's not. It makes it easier for candidates to deny real things that they said by suggesting that the content of e-mails and recordings were forged and that people shouldn't be believing their own eyes and ears. It's a net loss for our ability to form political consensus based on reality.

Meanwhile each state runs its own independent voter-registration system. Since 2016 many states have taken great strides to protect those systems by installing better network-intrusion detection systems or by upgrading antiquated hardware and software. But many have not.

During the last election, Russians probed or attempted to get into voter-registration systems in at least 18 states. Some sources quote higher numbers. And according to the Senate Select Committee on Intelligence's findings, in some of those states the Russians were in a position to alter or destroy the registration data. If they follow through this time, across entire states people will go to the polls and be told that they aren't on the lists. Maybe they will be given provisional ballots. But if this happens to a large fraction of voters, then there will be such terrible delays that many will give up and go home. A sophisticated attacker could even cause the registration system to lie to voters who confirm their own registration status through online portals while corrupting information in the rolls that are used in polling places.

Attacks on preelection functions could be engineered to have a racial or partisan effect. Because of antidiscrimination laws, some voter-registration records include not only political affiliation but also race. With access to that database, someone could easily manipulate only the records belonging to people of a certain political party, racial group or geographical location.

In some states, online voter-registration systems also allow the voter to request an absentee ballot or to change the address to which the ballot is directed. An attacker could request vote-by-mail ballots for a large number of citizens and direct them to people working with the attacker who would fill them in and cast fake votes.

ON ELECTION DAY

ELECTION INTERFERENCE can be successful in many ways—it depends on an attacker's goals and level of access. In a close election, if a coordinated group, say in Russia, thinks one candidate is much better than the other for their country, why not try to influence the outcome by undetectably manipulating votes? An attacker could infiltrate what are called election-management systems. There is a programming process by which the design of the ballot—the races and candidates and the rules for counting the votes—gets produced and then gets copied to every individual voting machine. Election officials usually copy it on memory cards or USB sticks for the election machines. That provides a route by which malicious code could spread from the centralized programming system to many voting machines in the field. Then the attack code runs on the individual voting machines, and it's just another piece of software. It has access to all the same data that the voting machine does, including all the electronic records of people's votes.

For 2020 I think ground zero for this kind of vote manipulation via cyberattack is an office building in the Midwest. Much of the country outsources its ballot design to just a few election vendors—the largest of which is a voting-machine manufacturer that, when I visited, told me it does the preelection programming for about 2,000 jurisdictions across 34 states. All of that's done from its headquarters, in a room I've been in that I'd describe as being part of a typical work building shared with other companies. If attackers can hack into that central facility and remotely infiltrate the company's computers, they can spread malicious code to voting machines and change election results across much of the country. The tactic might be as subtle as manipulating vote totals in close jurisdictions. It could easily go undetected.

The scientific consensus is that the best way to secure the vote is to use paper ballots and rigorously audit them, by having people inspect a random sample. Unfortunately, 12 states still don't have paper across the board. And some states, instead of adopting paper, are now having officials do auditing by looking at a scan of the original ballot on a computer screen. We have new research coming out that shows how you can use a computer algo-

© 2020 Scientific American

rithm to essentially do "deepfake" ballot scans. We used computer-vision techniques to automatically move the check marks around so that the scan of your ballot filled out in your distinctive handwriting reflects votes different from the ones you recorded on the piece of paper.

It might actually be scarier if attackers don't think one candidate is much better for their purposes than the other. Maybe their motivation is more general: to weaken American democracy. They could introduce malicious code that would make the election equipment essentially destroy itself when it is turned on in November 2020, which will cause massive chaos. Or they could have the equipment appear to work, but at the end of the day officials discover that no votes have been recorded. In the jurisdictions without paper backup, there is no other record of the vote. You would have to run a completely new election. The point of this kind of visible attack is that it undermines faith in the system and shakes people's confidence in the integrity of democracy.

ELECTION NIGHT AND BEYOND

YOU NEED TO GET PEOPLE to agree more or less about the truth and the conclusion of the election. But by the time November rolls around, we're all going to be primed to worry about the legitimacy of our process. So much is going to depend on how close the race seems on election night.

The way that results get transferred from your local precinct to the display on CNN or on the *New York Times* Web site is through a very centralized computer system operated by the Associated Press and others. What if an attacker were to hack those computer systems and cause the wrong call to be made on election night? We'd eventually find out about it because states go back and do their own totalization, but it might take days or even a couple of weeks until we discover a widespread error. People who want to believe the election was rigged would see this as confirmation it was rigged indeed.

Only 22 states have a requirement to complete any kind of post-election audit of their paper trail prior to legally certifying the results. And in 20 out of those 22 states, the requirement doesn't always result in a statistically significant level of auditing, because they do not look at a large enough ballot sample to have high confidence in the result, especially when results are close. It's just based on the math and has nothing to do with politics. Only Rhode Island and Colorado require a statistically rigorous process called a risk-limiting audit, although other states are moving in that direction.

If, because of computer hacking, we don't arrive at election results in many states, we enter unknown territory. The closest precedent would be something like the 2000 Bush versus Gore election where the outcome was ultimately decided in the Supreme Court and wasn't known for a month after election day. It would be terrifying, and it might involve running the election again in states that were affected. You really can't replay an election and expect to get the same results, because it's always going to be a different political environment.

Or let's say a candidate challenges a close election result. Under current rules and procedures, that is often the only way that people will ever go back and examine the physical evidence to check whether there was an attack. Right now we don't have the right forensic tools to be able to go back and see what happened where and who might have done what. It's not even clear who would have the jurisdiction to do those kinds of tests, because election officials and law enforcement don't often go hand in hand. You don't want to turn it over to the police to decide who won.

In a real nightmare scenario, attackers could gain enough access to the voting system to tip the election result and cause one candidate to win by fraud. Then they could keep that a secret—but engineer it in such a way that at any time in the future, they could prove they had stolen the election.

Imagine a swing state like Pennsylvania, which raced to replace its vulnerable paperless voting machines by the end of 2019. Despite this, the state still doesn't require risk-limiting audits, which means outcome-changing fraud could go undetected. What if the whole election comes down to Pennsylvania, and an attacker was able to hack into its machines and change the reported results? They could set the manipulation so that if you sorted the names of the polling places alphabetically, the least significant digits of the votes for the winning candidate formed the digits of pi—or something like that. It would be a pattern that wouldn't be noticeable but that could later be pointed in a way that undeniably shows the results were fake.

Say this information comes out after the new administration has been in power for a certain amount of time, and no one can deny that the president is not the legitimate winner. Now we have an unprecedented constitutional crisis. Finally, imagine if the nation-state that carries out this attack doesn't release its information publicly but instead uses it to blackmail the person who becomes president. This is pushing slightly into the realm of science fiction, though not by much.

The reality is that most cyberwarfare is more mundane. It's almost certain we're going to see attempts to sow doubt that are connected to the vulnerabilities in the election system just because it is so easy. You don't have to hack into a single piece of election equipment—all you have to do is suggest that someone might have.

It's hard to have an open conversation about the vulnerabilities in the system without risking contributing to attackers' goal of making people feel less confident in the results. But the fundamental problem is that the American election system is based on convincing the public to trust the integrity of the imperfect machinery and the imperfect people who operate it. Ultimately our best defense is to make elections be based on evidence instead of on faith—and it is entirely doable. There are so many problems in cybersecurity and critical infrastructure where you could offer me billions of dollars and decades to do research, and I'd say, *Maybe we can make this a little bit better.* But election-security challenges can be solved without any major scientific breakthroughs and for only a few hundred million dollars. It's just a matter of political will. SA

J. Alex Halderman is a professor of computer science and engineering at the University of Michigan, where he is also director of the Center for Computer Security and Society. He was a 2019 Andrew Carnegie Fellow for his work in educating lawmakers and the public in how to strengthen election cybersecurity.

Jen Schwartz is a senior editor at *Scientific American* for technology and mind.

MORE TO EXPLORE

Securing the Vote: Protecting American Democracy. National Academies of Sciences, Engineering, and Medicine. National Academies Press, 2018.

scientificamerican.com/magazine/sa

© 2020 Scientific American

THE TRIBALISM OF TRUTH

As political polarization grows, the arguments we have with one another may be shifting our understanding of truth itself

By Matthew Fisher, Joshua Knobe, Brent Strickland and Frank C. Keil

IN A KEY MOMENT OF THE FINAL TRUMP-CLINTON PRESIDENTIAL DEBATE, DONALD TRUMP turned to a question regarding Russian president Vladimir Putin:

"He has no respect for her," Trump said, pointing at Hillary Clinton. "Putin, from everything I see, has no respect for this person."

The two debaters then drilled down to try and gain a more nuanced understanding of the difficult policy issues involved. Clinton said,

"Are you suggesting that the aggressive approach I propose would actually fail to deter Russian expansionism?"

To which Trump responded,

"No, I certainly agree that it would deter Russian expansionism; it's just that it would also serve to destabilize the …"

Just kidding. That's not at all what happened. Actually each side aimed to attack and defeat the other. Clinton really said,

"Well, that's because he'd rather have a puppet as president of the United States."

To which Trump retorted,

"You're the puppet!"

IN BRIEF

The existence of moral objectivity is a thorny philosophical question. Cognitive scientists have gathered empirical evidence to see how ordinary people actually think about relativism versus immutable truth.

As political polarization grows, arguing to win is seemingly a more popular style of discourse than arguing to learn, especially in online forums such as Facebook and Twitter.

Researchers have found that the style of discourse people engage in actually changes their understanding of the question itself. If arguing to win is on the rise, it is very likely that objectivism is, too.

Illustration by Hanna Barczyk

© 2020 Scientific American

© 2020 Scientific American

Episodes like this one have become such a staple of contemporary political discourse that it is easy to forget how radically different they are from disputes we often have in ordinary life. Consider a couple of friends trying to decide on a restaurant for dinner. One might say, "Let's try the new Indian restaurant tonight. I haven't had Indian for months." To which another replies, "You know, I saw that place is getting poor reviews. Let's grab some pizza instead?" "Good to know—pizza it is," says the first. Each comes in with an opinion. They begin a discussion in which each presents an argument, then listens to the other's argument, and then they both move toward an agreement. This kind of dialogue happens all the time. In our research, which involves cognitive psychology and experimental philosophy, we refer to it as "arguing to learn."

But as political polarization increases in the U.S., the kind of antagonistic exchange exemplified by the Trump-Clinton debate is occurring with increasing frequency—not just among policy makers but among us all. In interactions such as these, people may provide arguments for their views, but neither side is genuinely interested in learning from the other. Instead the real aim is to "score points," in other words, to defeat the other side in a competitive activity. Conversations on Twitter, Facebook and even YouTube comment sections have become powerful symbols of what the combativeness of political discourse looks like these days. We refer to this kind of discussion as "arguing to win."

The divergence of Americans' ideology is accompanied by an animosity for those across the aisle. Recent polls show that partisan liberals and conservatives associate with one another less frequently, have unfavorable views of the opposing party, and would even be unhappy if a family member married someone from the other side. At the same time, the rise of social media has revolutionized how information is consumed—news is often personalized to one's political preferences. Rival perspectives can be completely shut out from one's self-created media bubble. Making matters worse, outrage-inducing content is more likely to spread on these platforms, creating a breeding ground for clickbait headlines and fake news. This toxic online environment is very likely driving Americans further apart and fostering unproductive exchanges.

In this time of rising tribalism, an important question has arisen about the psychological effects of arguing to win. What happens in our minds—and to our minds—when we find ourselves conversing in a way that simply aims to defeat an opponent? Our recent research has explored this question using experimental methods, and we have found that the distinction between different modes of argument has some surprisingly far-reaching effects. Not only does it change people's way of thinking about the debate and the people on the opposing side, but it also has a more fundamental effect on our way of understanding the very issue under discussion.

ARE WE OBJECTIVISTS OR RELATIVISTS?

THE QUESTION OF MORAL and political objectivity is a notoriously thorny one, which philosophers have been debating for millennia. Still, the core of the question is easy enough to grasp by considering a few hypothetical conversations. Consider a debate about a perfectly straightforward question in science or mathematics. Suppose two friends are working together on a problem and find themselves disagreeing about the solution:

Mary: The cube root of 2,197 is 13.
Susan: No, the cube root of 2,197 is 14.

People observing this conflict might not know which answer is correct. Yet they might be entirely sure that there is a single objectively correct answer. This is not just a matter of opinion—there is a fact of the matter, and anyone who has an alternative view is simply mistaken.

Now consider a different kind of scenario. Suppose these two friends decide to take a break for lunch and find themselves disagreeing about what to put on their bagels:

Mary: Veggie cream cheese is really tasty.
Susan: No, veggie cream cheese is not tasty at all. It is completely disgusting.

In this example, observers might take up another attitude: Even if two people have opposite opinions, it could be that neither is incorrect. It seems that there is no objective truth of the matter.

With that in mind, think about what happens when people debate controversial questions about morally infused political topics. As our two friends are enjoying their lunch, suppose they wade into a heated political chat:

Mary: Abortion is morally wrong and should not be legal.
Susan: No, there is nothing wrong with abortion, and it should be perfectly legal.

The question we grapple with is how to understand this kind of debate. Is it like the math question, where there is an objectively right answer and anyone who says otherwise must be mistaken? Or is it more like a clash over a matter of taste, where there is no single right answer and people can have opposite opinions without either one being wrong?

In recent years work on this topic has expanded beyond the realm of philosophy and into psychology and cognitive science. Instead of relying on the intuitions of professional philosophers, researchers like ourselves have begun gathering empirical evidence to understand how people actually think about these issues. Do people tend to think moral and political questions have objectively correct answers? Or do they have a more relativist view?

On the most basic level, the past decade of research has shown that the answer to this question is that it's complicated. Some people are more objectivist; others are more relativist. That might seem obvious, but later studies explored the differences between people with these types of thinking. When participants are asked whether they would be willing to share an apartment with a roommate who holds opposing views on moral or political questions, objectivists are more inclined to say no. When participants are asked to sit down in a room next to a person who has opposing views, objectivists actually sit farther away. As University of Pennsylvania psychologist Geoffrey P. Goodwin once put it, people who hold an objectivist view tend to respond in a more "closed" fashion.

Why might this be? One straightforward possibility is that if you think there is an objectively correct answer, you may be drawn to conclude that everyone who holds the opposite view is

© 2020 Scientific American

simply incorrect and therefore not worth listening to. Thus, people's view about objective moral truths could shape their approach to interacting with others. This is a plausible hypothesis and one worth investigating in further studies. Yet we thought that there might be more to the story. In particular, we suspected there might be an effect in the opposite direction. Perhaps it's not just that having objectivist views shapes your interactions with other people; perhaps your interactions with other people can actually shape the degree to which you hold objectivist views.

WINNING VS. LEARNING

TO TEST THIS THEORY, we ran an experiment in which adults engaged in an online political conversation. Each participant logged on to a Web site and indicated his or her positions on a variety of controversial political topics, including abortion and gun rights. They were matched with another participant who held opposing views. The participants then engaged in an online conversation about a topic on which they disagreed.

Half of the participants were encouraged to argue to win. They were told that this would be a highly competitive exchange and that their goal should be to outperform the other person. The result was exactly the kind of communication one sees every day on social media. Here, for example, is a transcript from one of the actual conversations:

P1: I believe 100 percent in a woman's choice.
P2: Abortion should be prohibited because it stops a beating heart.
P1: Abortion is the law of the land, the land you live in.
P2: The heart beats at 21 days its murder [sic].

The other half of participants were encouraged to argue to learn. They were told that this would be a very cooperative exchange and that they should try to learn as much as they could from their opponent. These conversations tended to have a quite different tone:

P3: I believe abortion is a right all women should possess. I do understand that some people choose to place certain determinants on when and why, but I think it should be for any reason before a certain time point in the pregnancy agreed upon by doctors, so as not to harm the mother.
P4: I believe that life begins at conception (sperm meeting egg), so abortion to me is the equivalent of murder.
P3: I can absolutely see that point. As a biologist, it is obvious from the first cell division that "life" is happening. But I do not think life is advanced enough to warrant abolishing abortion.

It is not all that surprising that these two sets of instructions led to such results. But would these exchanges in turn lead to different views about the very nature of the question being discussed? After the conversation was over, we asked participants whether they thought there was an objective truth about the topics they had just debated. Strikingly, these 15-minute exchanges actually shifted people's views. Individuals were more objectivist after arguing to win than they were after arguing to learn. In other words, the social context of the discussion—how people frame the purpose of controversial discourse—actually changed their opinions on the deeply philosophical question about whether there is an objective truth at all.

These results naturally lead to another question that goes beyond what can be addressed through a scientific study. Which of these two modes of argument would be better to adopt when it comes to controversial political topics? At first, the answer seems straightforward. Who could fail to see that there is something deeply important about cooperative dialogue and something fundamentally counterproductive about sheer competition?

Although this simple answer may be right most of the time, there may also be cases in which things are not quite so clear-cut. Suppose we are engaged in a debate with a group of climate science skeptics. We could try to sit down together, listen to the arguments of the skeptics and do our best to learn from everything they have to say. But some might think that this approach is exactly the wrong one. There might not be anything to be gained by remaining open to ideas that contradict scientific consensus. Indeed, agreeing to partake in a cooperative dialogue might be an instance of what journalists call "false balance"—legitimizing an extreme outlier position that should not be weighed equally. Some would say that the best approach in this kind of case is to argue to win.

Of course, our studies cannot directly determine which mode of argument is "best." And although plenty of evidence suggests that contemporary political discourse is becoming more combative and focused on winning, our findings do not elucidate *why* that change has occurred. Rather they provide an important new piece of information to consider: the mode of argument we engage in actually changes our understanding of the question itself. The more we argue to win, the more we will feel that there is a single objectively correct answer and that all other answers are mistaken. Conversely, the more we argue to learn, the more we will feel that there is no single objective truth and different answers can be equally right.

So the next time you are deciding how to enter into an argument on Facebook about the controversial question of the day, remember that you are not just making a choice about how to interact with a person who holds the opposing view. You are also making a decision that will shape the way you—and others—think about whether the question itself has a correct answer. SA

Matthew Fisher is an assistant professor of psychology at Southern Methodist University.

Joshua Knobe is a professor at Yale University, appointed both in the program in cognitive science and in the department of philosophy.

Brent Strickland is a researcher in cognitive science at the Jean Nicod Institute in Paris.

Frank C. Keil is Charles C. and Dorathea S. Dilley Professor of Psychology and a professor of linguistics and cognitive science at Yale University.

MORE TO EXPLORE

Why Are Some Moral Beliefs Perceived to Be More Objective Than Others? G. P. Goodwin and J. M. Darley in *Journal of Experimental Social Psychology*, Vol. 48, No. 1, pages 250–256; January 2012.
The Influence of Social Interaction on Intuitions of Objectivity and Subjectivity. Matthew Fisher et al. in *Cognitive Science*, Vol. 41, No. 4, pages 1119–1134; May 2017.

scientificamerican.com/magazine/sa

© 2020 Scientific American

Post-Truth: A Guide for the Perplexed

If politicians can lie without condemnation, what are scientists to do?

By Kathleen Higgins

THE OXFORD DICTIONARIES NAMED "POST-TRUTH" AS THEIR 2016 WORD of the Year. It must sound alien to scientists. Science's quest for knowledge about reality presupposes the importance of truth, both as an end in itself and as a means of resolving problems. How could truth become passé? For philosophers like me, post-truth also goes against the grain. But in the wake of the 2016 U.S. presidential election and the seemingly endless campaigns preceding it, author Ralph Keyes's 2004 declaration that we have arrived in a post-truth era seems distressingly plausible.

Post-truth refers to blatant lies being routine across society, and it means that politicians can lie without condemnation. This is different from the cliché that all politicians lie and make promises they have no intention of keeping—this still expects honesty to be the default position. In a post-truth world, this expectation no longer holds.

This can explain the current political situation in the U.S. and elsewhere. Public tolerance of inaccurate and undefended allegations, non sequiturs in response to hard questions and outright denials of facts is shockingly high. Repetition of talking points passes for political discussion, and serious interest in issues and options is treated as the idiosyncrasy of wonks. The lack of public indignation when political figures claim disbelief in response to growing scientific evidence of the reality of climate change is part of this larger pattern. "Don't bother me with facts" is no longer a punchline. It has become a political stance. It's worth remembering that it has not always been this way: the exposure of former U.S. president Richard Nixon's lies was greeted with outrage.

One might be tempted to blame philosophy for post-truth. Some of us write about epistemic relativism, the view that truth can vary depending on the context. Yet relativism is itself relative. An extreme relativist might hold that the truth varies from person to person, a position that does not leave much room for debate. But more rational positions can also involve at least a modicum of relativism. In a sense, even 18th-century philosopher Immanuel Kant's quite sensible contention that we can never know what things are like "in themselves"—independent of how our minds format what we perceive—is a relativistic position. It implies that what is true of the world for humans is probably different from what is true for a fly. Entomologists would surely agree.

More radical forms of relativism are often denounced as undermining basic values. Friedrich Nietzsche, the 19th-century philosopher who is often invoked to justify post-truth, was such a relativist, and he does suggest at times that deception is rife and should not be categorically rejected. His point is to complicate our view of human behavior and to object to moral certainties that encourage black-and-white judgments about what's good and what's evil. Thus, he denies that there are moral facts, saying that we have only "moral interpretations," and in doing so he denies that moral assertions are unconditionally true. But this does not mean there is no truth. Even when he claims that our truths amount to our "irrefutable errors," he is pointing to the exaggerated clarity of abstractions by comparison with empirical reality.

In fact—contrary to how he is often presented—Nietzsche held intellectual honesty at a premium. His most strenuous rejections of "truth" are mostly directed not at truth but at what has been asserted as true. Yes, Nietzsche was an elitist who was skeptical of democracy, and so his work does not necessarily fault leaders for talking down to the public. But it also points out the inconsistency of re-

© 2020 Scientific American

ligious teachers who assume they have the right to lie.

When political leaders make no effort to ensure that their "facts" will bear scrutiny, we can only conclude that they take an arrogant view of the public. They take their right to lie as given, perhaps particularly when the lies are transparent. Many among the electorate seem not to register the contempt involved, perhaps because they would like to think that their favored candidate is at least well intentioned and would not deliberately mislead them.

Much of the public hears what it wants to hear because many people get their news exclusively from sources whose bias they agree with. But contemptuous leaders and voters who are content with handwaving and entertaining bluster undermine the democratic idea of rule by the people. The irony is that politicians who benefit from post-truth tendencies rely on truth, too, but not because they adhere to it. They depend on most people's good-natured tendency to trust that others are telling the truth, at least the vast majority of the time.

Scientists and philosophers should be shocked by the idea of post-truth, and they should speak up when scientific findings are ignored by those in power or treated as mere matters of faith. Scientists must keep reminding society of the importance of the social mission of science—to provide the best information possible as the basis for public policy. And they should publicly affirm the intellectual virtues that they so effectively model: critical thinking, sustained inquiry and revision of beliefs on the basis of evidence. Another line from Nietzsche is especially pertinent now: "Three cheers for physics!—and even more for the motive that spurs us toward physics—our honesty!" SA

Kathleen Higgins teaches and writes on Nietzsche, aesthetics and philosophy of emotion at the University of Texas at Austin.

Illustration by Hannah Salyer

© 2020 Scientific American

Why People "Fly from Facts"

Research shows the appeal of untestable beliefs and how they lead to a polarized society

By Troy Campbell and Justin Friesen

THERE WAS A SCIENTIFIC STUDY THAT SHOWED vaccines cause autism." "Actually, the researcher in that study lost his medical license, and overwhelming research since then has shown no link between vaccines and autism." "Well, regardless, it's still my personal right as a parent to make decisions for my child." Does that exchange sound familiar? A debate that starts with testable factual statements, but then, when the truth becomes inconvenient, the person takes a flight from facts.

As public debate rages about issues such as immunization, Obamacare and same-sex marriage, many people try to use science to bolster their arguments. And because it's becoming easier to test and establish facts—whether in physics, psychology or policy—many have wondered why bias and polarization have not been defeated. When people are confronted with facts, such as the well-established safety of immunization, why do these facts seem to have so little effect?

Our research, published in 2015 in the *Journal of Personality and Social Psychology,* examined a slippery way by which people get away from facts that contradict their beliefs. Of course, sometimes people just dispute the validity of specific facts. But we find that sometimes they go one step further and, as in the opening example, reframe an issue in untestable ways. This makes potential important facts and science ultimately irrelevant to the issue.

Let's consider the issue of same-sex marriage. Facts could be relevant to whether it should be legal—for example, if data showed that children raised by same-sex parents are worse off—or just as well off—as children raised by opposite-sex parents. But what if those facts contradict one's views?

We presented 174 American participants who supported or opposed same-sex marriage with (supposed) scientific facts that supported or disputed their position. When the facts opposed their views, our participants—on both sides of the issue—were more likely to state that same-sex marriage isn't actually about facts: it's more a question of moral opinion. But when the facts were on their side, they more often stated that their opinions were fact-based and much less about morals. In other words, we observed something beyond the denial of particular facts: we observed a denial of the relevance of facts.

In a similar study using 117 religious participants, we had some read an article critical of religion. Believers who were especially high (but not low) in religiosity were more likely to turn to more untestable "blind faith" arguments as reasons for their beliefs than to arguments based in factual evidence, compared with those who read a neutral article.

These experiments show that when people's beliefs are threatened, they often take flight to a land where facts do not matter. In scientific terms, their beliefs become less "falsifiable" because they can no longer be tested scientifically for verification or refutation.

For instance, sometimes people dispute government policies based on the argument that they don't work. Yet if facts suggest that the policies do work, the same person might stay resolvedly against the argument based on principle. We can see this on both sides of the political spectrum, whether it's conservatives and Obamacare or liberals and the "surge" in Iraq in 2007.

One would hope that objective facts could allow people to reach consensus more easily, but American politics are more polarized than ever. Could this polarization be a consequence of feeling free of facts?

While it is difficult to objectively test that idea, we can experimentally assess a fundamental question: When people are made to see their important beliefs as relatively less rather than more testable, does it increase polarization and commitment to desired beliefs? Two experiments we conducted suggest so.

In an experiment with 179 Americans (conducted during President Barack Obama's tenure), we reminded roughly half of participants that much of Obama's policy performance was empirically testable and did not remind the other half. Then participants rated Obama's performance on five domains (such as job creation). Comparing opponents and supporters of Obama, we found that the reminder of testability reduced the average polarized assessments of Obama's performance by about 40 percent.

To further test the hypothesis that people strengthen their desired beliefs when the beliefs are free of facts, we looked at a sample of 103 participants that varied from highly to moderately religious. We found that when highly (but not more moderately) religious participants were told that God's existence will always be untestable, they reported stronger desirable religious beliefs afterward (such as the be-

© 2020 Scientific American

lief that God was looking out for them), relative to when they were told that one day science might be able to investigate God's existence.

Together these findings show, at least in some cases, when testable facts are less a part of the discussion, people dig deeper into the beliefs they wish to have—such as viewing a politician in a certain way or believing God is constantly there to provide support. These results bear similarities to the many studies that find when facts are fuzzier, people tend to exaggerate desired beliefs.

So after examining the power of untestable beliefs, what have we learned about dealing with human psychology? We have learned that bias is a disease and to fight it we need a healthy treatment of facts and education. We find that when facts are injected into the conversation, the symptoms of bias become less severe. But, unfortunately, we have also learned that facts can only do so much. To avoid coming to undesirable conclusions, people can fly from the facts and use other tools in their deep, belief-protecting toolbox.

With the disease of bias, then, societal immunity is better achieved when people are encouraged to accept ambiguity, engage in critical thinking and reject strict ideology. This society is something the Common Core State Standards for education and at times *The Daily Show* are at least in theory helping to create. We will never eradicate bias—not from others, not from ourselves, not from society. But we can become more free of ideology and less free of facts. SA

Troy Campbell is a freelance consulting psychologist for business. He has a doctoral degree from Duke University. He studies how our beliefs and identities centrally affect everyday experiences and thought.

Justin Friesen is an assistant professor of psychology at the University of Winnipeg. He studies the psychological and motivational factors that perpetuate social and economic inequality.

Illustration by Thomas Fuchs

© 2020 Scientific American

WHY WE BELIEVE CONSPIRACY

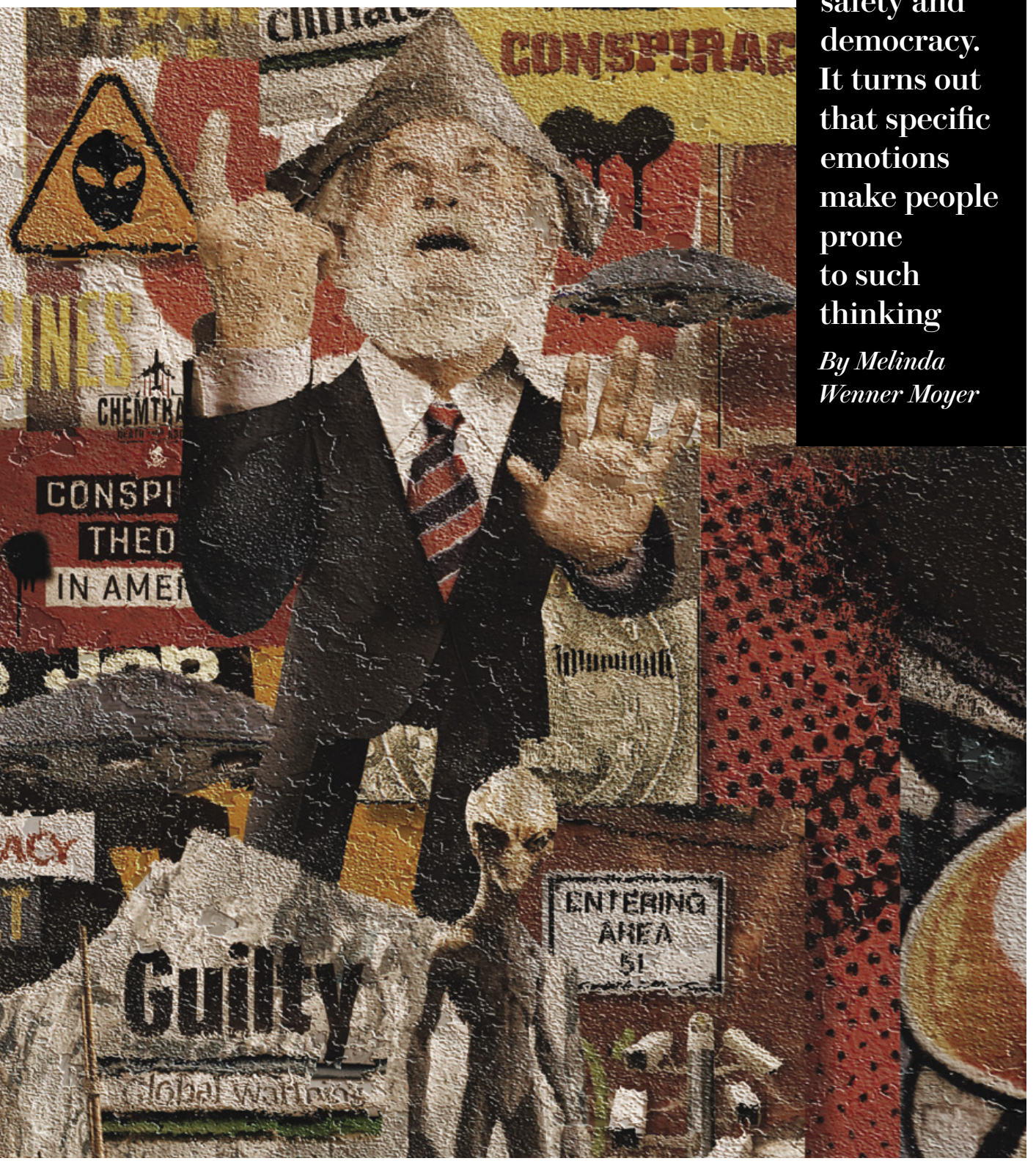
Illustration by Eddie Guy

© 2020 Scientific American

THEORIES

Baseless theories threaten our safety and democracy. It turns out that specific emotions make people prone to such thinking

By Melinda Wenner Moyer



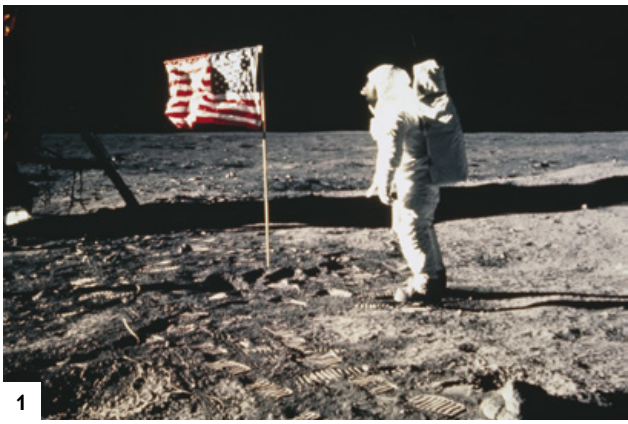
© 2020 Scientific American

STEPHAN LEWANDOWSKY WAS DEEP IN DENIAL. ABOUT SIX years ago the cognitive scientist had thrown himself into a study of why some people refuse to accept the overwhelming evidence that the planet is warming and humans are responsible. As he delved into this climate change denialism, Lewandowsky, then at the University of Western Australia, discovered that many of the naysayers also believed in outlandish plots, such as the idea that the *Apollo* moon landing was a hoax created by the American government. "A lot of the discourse these people were engaging in on the Internet was totally conspiratorial," he recalls.

Lewandowsky's findings, published in 2013 in *Psychological Science,* brought these conspiracy theorists out of the woodwork. Offended by his claims, they criticized his integrity online and demanded that he be fired. (He was not, although he has since moved to the University of Bristol in England.) But as Lewandowsky waded through one irate post after another, he discovered that his critics—in response to his assertions about their conspiratorial tendencies—were actually spreading new conspiracy theories *about him*. These people accused him and his colleagues of faking survey responses and of conducting the research without ethical approval. When his personal Web site crashed, one blogger accused him of intentionally blocking critics from seeing it. None of it was true.

The irony was amusing at first, but the ranting even included a death threat, and calls and e-mails to his university became so vicious that the administrative staff who fielded them asked their managers for help. That was when Lewandowsky changed his assessment. "I quickly realized that there was nothing funny about these guys at all," he says.

The dangerous consequences of the conspiratorial perspective—the idea that people or groups are colluding in hidden ways to produce a particular outcome—have become painfully clear. The belief that the coronavirus pandemic is an elaborate hoax designed to prevent the reelection of Donald Trump has incited some Americans to forgo important public health recommendations, costing lives. The gunman who shot and killed 11 people and injured six others in a Pittsburgh synagogue in October 2018 justified his attack by claiming that Jewish people were stealthily supporting illegal immigrants. In 2016 a conspiracy theory positing that high-ranking Democratic Party officials were involved in a child sex ring involving several Washington, D.C., area restaurants incited one believer to fire an assault weapon inside a pizzeria. Luckily no one was hurt.


1

The mindset is surprisingly common, although thankfully it does not often lead to gunfire. More than a quarter of the American population believes there are conspiracies "behind many things in the world," according to a 2017 analysis of government survey data by University of Oxford and University of Liverpool researchers. The prevalence of conspiracy mongering may not be new, but today the theories are

PRECEDING PAGES: GETTY IMAGES (*illustration reference*); THIS PAGE: GETTY IMAGES (1)

IN BRIEF

False conspiracy theories can drive people to violence, as they did for the Pittsburgh synagogue shooter, and affect political activity.

Anxious people are especially drawn to conspiratorial thinking, experiments show, and the mindset is also triggered by a loss of control.

You can spot hallmarks of fake theories, such as internal contradictions in the "evidence" and contentions based on shaky assumptions, psychologists say.

© 2020 Scientific American

becoming more visible, says Viren Swami, a social psychologist at Anglia Ruskin University in England, who studies the phenomenon. For instance, when more than a dozen bombs were sent to prominent Democrats and Trump critics, as well as CNN, in October 2018, a number of high-profile conservatives quickly suggested that the explosives were really a "false flag," a fake attack orchestrated by Democrats to mobilize their supporters during the U.S. midterm elections.

One obvious reason for the current raised profile of this kind of thinking is that the U.S. president is a vocal conspiracy theorist. Donald Trump has suggested, among other things, that the father of Senator Ted Cruz of Texas helped to assassinate President John F. Kennedy and that Democrats funded the same migrant caravan traveling from Honduras to the U.S. that worried the Pittsburgh synagogue shooter.

is now supported by a slew of guilty pleas, evidence-based indictments and U.S. intelligence agency conclusions. So how is one to know what to believe? There, too, psychologists have been at work and have uncovered strategies that can help people distinguish plausible theories from those that are almost certainly fake—strategies that seem to become more important by the day.

THE ANXIETY CONNECTION

IN MAY 2018 the American Psychiatric Association released the results of a national survey suggesting that 39 percent of Americans feel more anxious than they did a year ago, primarily about health, safety, finances, politics and relationships. Another 2017 report found that 63 percent of Americans are extremely worried about the future of the nation and that 59 percent consider this the lowest point in U.S. history that they can remember. These feelings span the political spectrum. A 2018 Pew Research Center survey found that the majority of both Democrats and Republicans feel that "their side" in politics has been losing in recent years on issues they find important.

2

3

4

CONSPIRACY THEORISTS believe plots are behind many situations. Some hold that the *Apollo* moon landing was faked (*1*), others that the White House forced Supreme Court Justice Anthony Kennedy to retire (*2*), and others that Trump slogans on a mail bomber's van were put there to frame Republicans (*3*). The gunman who killed 11 synagogue members in 2018 claimed a Jewish group was undermining America (*4*).

BRENDAN SMIALOWSKI *Getty Images (2)*; PAUL BILODEAU *Alamy (3)*; JEFF SWENSEN *Getty Images (4)*

But there are other factors at play, too. New research suggests that events happening worldwide are nurturing underlying emotions that make people more willing to believe in conspiracies. Experiments have revealed that feelings of anxiety make people think more conspiratorially. Such feelings, along with a sense of disenfranchisement, currently grip many Americans, according to surveys. In such situations, a conspiracy theory can provide comfort by identifying a convenient scapegoat and thereby making the world seem more straightforward and controllable. "People can assume that if these bad guys weren't there, then everything would be fine," Lewandowsky says. "Whereas if you don't believe in a conspiracy theory, then you just have to say terrible things happen randomly."

Discerning fact from fiction can be difficult, however, and some seemingly wild conspiracy ideas turn out to be true. The once scoffed at notion that Russian nationals meddled in the 2016 presidential election

Such existential crises can promote conspiratorial thinking. In a 2015 study in the Netherlands, researchers split college students into three groups. People in one group were primed to feel powerless. The scientists asked them to recall and write about a time in their lives when they felt they were not in control of the situation they were in. Those in a second group were cued in the opposite direction. They were asked to write about a time when they felt totally in control. And still others, in a third group, were asked something neutral: to describe what they had for dinner last night. Then the researchers asked all the groups how they felt about the construction of a new subway line in Amsterdam that had been plagued by problems.

© 2020 Scientific American

Students who had been primed to feel in control were less likely than students in the other two groups to support conspiracy theories regarding the subway line, such as the belief that the city council was stealing from the subway's budget and that it was intentionally jeopardizing residents' safety. Other studies have uncovered similar effects. Swami and his colleagues, for instance, reported in 2016 that individuals who feel stressed are more likely than others to believe in conspiracy theories, and a 2017 study found that promoting anxiety in people also makes them more conspiracy-minded.

Feeling alienated or unwanted also seems to make conspiratorial thinking more attractive. In 2017 Princeton University psychologists set up an experiment with trios of people. The researchers asked all participants to write two paragraphs describing themselves and then told them that their descriptions would be shared with the other two in their group, who would use that information to decide if they would work with the person in the future. After telling some subjects that they had been accepted by their group and others that they had been rejected, the researchers evaluated the subjects' thoughts on various conspiracy-related scenarios. The "rejected" participants, feeling alienated, were more likely than the others to think the scenarios involved a coordinated conspiracy.

When feelings of personal alienation or anxiety are combined with a sense that society is in jeopardy, people experience a kind of conspiratorial double whammy, according to a study conducted near the start of the U.S.'s Great Recession.

It is not just personal crises that encourage individuals to form conspiratorial suspicions. Collective social setbacks do so as well. In a 2018 study, researchers at the University of Minnesota and Lehigh University surveyed more than 3,000 Americans. They found that participants who felt that American values are eroding were more likely than others to agree with conspiratorial statements, such as that "many major events have behind them the actions of a small group of influential people." Joseph Uscinski, a political scientist at the University of Miami, and his colleagues have shown that people who dislike the current political party in power think more conspiratorially than those who support the controlling party. Recently in the U.S., a number of unproved conjectures have come from political liberals as conservatives have ascended to control the government. These include the charge that the White House coerced Anthony Kennedy to retire from the U.S. Supreme Court and the allegation that Russian president Vladimir Putin is blackmailing Trump with a video of him watching prostitutes urinate on a Moscow hotel bed.

When feelings of personal alienation or anxiety are combined with a sense that society is in jeopardy, people experience a kind of conspiratorial double whammy. In a study conducted in 2009, near the start of the U.S.'s Great Recession, Daniel Sullivan, a psychologist now at the University of Arizona, and his colleagues told one group that parts of their lives were largely out of their control because they could be exposed to a natural disaster or some other catastrophe and told another group that things were under their control. Then participants were asked to read essays that argued that the government was handling the economic crisis either well or poorly. Those cued about uncontrolled life situations and told their government was doing a bad job were the most likely to think that negative events in their lives would be instigated by enemies rather than random chance, which is a conspiratorial hallmark.

While humans seek solace in conspiracy theories, however, they rarely find it. "They're appealing but not necessarily satisfying," says Daniel Jolley, a psychologist at Staffordshire University in England. For one thing, conspiratorial thinking can incite individuals to behave in a way that increases their sense of powerlessness, making them feel even worse. A 2014 study co-authored by Jolley found that people who are presented with conspiracy theories about climate change—scientists are just chasing grant money, for instance—are less likely to plan to vote, whereas a 2017 study reported that believing in work-related conspiracies—such as the idea that managers make decisions to protect their own interests—causes individuals to feel less committed to their job. "It can snowball and become a pretty vicious, nasty cycle of inaction and negative behavior," says Karen Douglas, a psychologist at the University of Kent in England and a co-author of the paper on work-related conspiracies.

The negative and alienated beliefs can also promote dangerous behaviors in some, as with the Pittsburgh shootings and the pizzeria attack. But the theories need not involve weapons to inflict harm. People who believe vaccine conspiracy theories, for example, say they are less inclined to vaccinate their kids, which creates pockets of infectious disease that put entire communities at risk.

TELLING FACT FROM FICTION

IT MAY BE POSSIBLE to quell conspiracy ideation, at least to some degree. One long-standing question has been whether or not it is a good idea to counter con-

© 2020 Scientific American

spiracy theories with logic and evidence. Some older research has pointed to a "backfire effect"—the idea that refuting misinformation can just make individuals dig their heels in deeper. "If you think there are powerful forces trying to conspire and cover [things] up, when you're given what you see as a cover story, it only shows you how right you are," Uscinski says.

But more recent research suggests that this putative effect is, in fact, rare. A 2016 study reported that when researchers refuted a conspiracy theory by pointing out its logical inconsistencies, it became less enchanting to people. And in a paper published online in 2018 in *Political Behavior,* researchers recruited more than 10,000 people and presented them with corrections to various claims made by political figures. The authors concluded that "evidence of factual backfire is far more tenuous than prior research suggests." In a recent review, the researchers who first described the backfire effect said that it may arise most often when people are being challenged over ideas that define their worldview or sense of self. Finding ways to counter conspiracy theories without challenging a person's identity may therefore be an effective strategy.

Encouraging analytic thinking may also help. In a 2014 study published in *Cognition*, Swami and his colleagues recruited 112 people for an experiment. First, they had everyone fill out a questionnaire that evaluated how strongly they believed in various conspiracy theories. A few weeks later the subjects came back in, and the researchers split them into two groups. One group completed a task that included unscrambling words in sentences containing words such as "analyze" and "rational," which primed them to think more analytically. The second group completed a neutral task.

Then the researchers readministered the conspiracy theory test to the two groups. Although the groups had been no different in terms of conspiratorial thinking at the beginning of the experiment, the subjects who had been incited to think analytically became less conspiratorial. Thus, by giving people "the tools and the skills to analyze data and to look at data critically and objectively," we might be able to suppress conspiratorial thinking, Swami says.

Analytic thinking can also help discern implausible theories from ones that, crazy as they sound, are supported by evidence. Karen Murphy, an educational psychologist at Pennsylvania State University, suggests that individuals who want to improve their analytic thinking skills should ask three key questions when interpreting conspiracy claims. One: What is your evidence? Two: What is your source for that evidence? Three: What is the reasoning that links your evidence back to the claim? Sources of evidence need to be accurate, credible and relevant. For instance, "you shouldn't take advice from your mom about whether the yellow color under your fingernails is a bad sign," Murphy says—that kind of information should come from someone who has expertise on the topic, such as a physician.

In addition, false conspiracy theories have several hallmarks, Lewandowsky says. Three of them are particularly noticeable. First, the theories include contradictions. For example, some deniers of climate change argue that there is no scientific consensus on the issue while framing themselves as heroes pushing back against established consensus. Both cannot be true. A second telltale sign is when a contention is based on shaky assumptions. Trump, for instance, claimed that millions of illegal immigrants cast ballots in the 2016 presidential election and were the reason he lost the popular vote. Beyond the complete lack of evidence for such voting, his assumption was that multitudes of such votes—if they existed—would have been for his Democratic opponent. Yet past polls of unauthorized Hispanic immigrants suggest that many of them would have voted for a Republican candidate over a Democratic one.

A third sign that a claim is a far-fetched theory, rather than an actual conspiracy, is that those who support it interpret evidence against their theory as evidence *for* it. When the van of the convicted mail bomber Cesar Sayoc was found in Florida plastered with Trump stickers, for instance, some individuals said this helped to prove that Democrats were really behind the bombs. "If anyone thinks this is what a real conservative's van looks like, you are being willfully ignorant. Cesar Sayoc is clearly just a fall guy for this obvious false flag," one person posted on Twitter.

Conspiracy theories are a human reaction to confusing times. "We're all just trying to understand the world and what's happening in it," says Rob Brotherton, a psychologist at Barnard College and author of *Suspicious Minds: Why We Believe in Conspiracy Theories* (Bloomsbury Sigma, 2015). But real harm can come from such thinking, especially when believers engage in violence as a show of support. By looking out for suspicious signatures and asking thoughtful questions about the stories we encounter, it is still possible to separate truth from lies. It may not always be an easy task, but it is a crucial one for all of us. SA

Melinda Wenner Moyer is a contributing editor at *Scientific American*. She wrote about multidisease vaccines in the June 2019 issue.

MORE TO EXPLORE

NASA Faked the Moon Landing—Therefore, (Climate) Science Is a Hoax: An Anatomy of the Motivated Rejection of Science. Stephan Lewandowsky et al. in *Psychological Science,* Vol. 24, No. 5, pages 622–633; May 2013.

The Influence of Control on Belief in Conspiracy Theories: Conceptual and Applied Extensions. Jan-Willem van Prooijen and Michele Acker in *Applied Cognitive Psychology,* Vol. 29, No. 5, pages 753–761; September/October 2015.

Putting the Stress on Conspiracy Theories: Examining Associations between Psychological Stress, Anxiety, and Belief in Conspiracy Theories. Viren Swami et al. in *Personality and Individual Differences,* Vol. 99, pages 72–76; September 2016.

Suspicion in the Workplace: Organizational Conspiracy Theories and Work-Related Outcomes. Karen M. Douglas and Ana C. Leite in *British Journal of Psychology,* Vol. 108, No. 3; pages 486–506; August 2017.

scientificamerican.com/magazine/sa

© 2020 Scientific American

© 2020 Scientific American

CONTAGIOUS DISHONESTY

Dishonesty begets dishonesty, rapidly spreading unethical behavior through a society

By Dan Ariely and Ximena Garcia-Rada

Imagine that you go to city hall for a construction permit to renovate your house. The employee who receives your form says that, because of the great number of applications the office has received, the staff will take up to nine months to issue the permit. But if you give her $100, your form will make it to the top of the pile. You realize she has just asked for a bribe: an illicit payment to obtain preferential treatment. A number of questions are likely to go through your head. Will I pay to speed things up? Would any of my friends or relatives do the same? You would probably not wonder, however, whether being exposed to the request would, in and of itself, affect a subsequent ethical decision. That is the kind of question behavioral researchers ask to investigate how corruption spreads.

The extent of bribery is hard to measure, but estimates from the World Bank suggest that corrupt exchanges involve $1 trillion annually. In 2018 Transparency International reported that more than two thirds of 180 countries it surveyed got a score of less than 50 on a scale from 0 ("highly corrupt") to 100 ("very clean"). Major scandals regularly make global headlines, such as when Brazilian construction company Odebrecht admitted in 2016 to having paid upward of $700 million in bribes to politicians and bureaucrats in 12 countries. But petty corruption, involving small favors between a few people, is also very common. Transparency International's Global Corruption Barometer for 2017 shows that one in every four of those surveyed said they had paid a bribe when accessing public services in the previous year, with almost one in three reporting such payments in the Middle East and North Africa.

Corruption, big or small, impedes the socioeconomic development of nations. It affects economic activities, weakens institutions, interferes with democracy and erodes the public's trust in government officials, politicians and their neighbors. Understanding the underlying psychology of bribery could be crucial to tackling the problem. Troublingly, our studies suggest that mere exposure to corruption is corrupting. Unless preventive measures are taken, dishonesty can spread stealthily and uninvited from person to person like a disease, eroding social norms and ethics—and once a culture of cheating and lying becomes entrenched, it can be difficult to dislodge.

CONTAGION

Suppose you refused the City Hall employee's request for a bribe. How would the experience influence your response to a subsequent ethical dilemma? In laboratory studies we conducted with behavioral researchers Vladimir Chituc, Aaron Nichols, Heather Mann, Troy Campbell and Panagiotis Mitkidis, we sought an answer to that question.

We invited individuals to the behavioral lab in the university to play a game that involved throwing a virtual die for a reward. Everyone was told that they would be compensated based on the outcome of multiple rolls. In practice, however, they could misreport their rolls to earn more money. So all participants faced a conflict between playing the game by the rules and behaving dishonestly to earn more. We created this setup to assess how individuals balance external and internal—or psychological—rewards when making ethical decisions. Research that Nina Mazar, On Amir and one of us (Ariely) published in 2008 indicates that most people act unethically to the extent that they can benefit while also preserving their moral self-image—an observation they described as the theory of self-concept maintenance.

Our game involved rolling a virtual die 30 times on iPads. Many behavioral economists have used similar paradigms involving physical dice and coins to assess dishonesty in so-called decontextualized games—that is, games that are not affected by social or cultural norms. Prior to each roll, participants were instructed to choose a side of the die in their mind—top or bottom—and report their choice *after* seeing the outcome of the roll. They would earn a fixed amount of money per dot on the side they reported each time. So everyone had a financial incentive to cheat by reporting the high-paying side. For example, if the outcome of the roll was two on the top of the die and five on the bottom of the die, people might be tempted to say they had chosen "bottom" before the roll even if they had not.

This paradigm does not allow us to know whether someone cheated in a specific roll. Nevertheless, when results are aggregated across all rolls and participants in a group, the proportion of favorable rolls chosen can be compared against chance (50 percent) to assess the magnitude of dishonesty. After participants received instructions about the game and how they would make money in the session, which they would get to take home, they were randomly assigned to a low- or a high-payment version. Those in the high-

IN BRIEF

Corruption damages economies, institutions and democratic structures.
Exposure to bribery can, in and of itself, be corrupting—suggesting a mechanism by which unethical behavior may spread through society.
Social norms influence ethical behavior. Surprisingly, however, the innate tendency to cheat (or not) is the same across countries, despite vast differences in actual corruption levels.
Further research is required into what drives bribery and corruption, how it spreads and how it can be controlled.

Illustration by Lisk Feng

© 2020 Scientific American

payment game would do exactly the same as those in the low-payment game but earn 10 times more. Everyone was told about the existence of the other game. Then, half the participants in the low-payment condition were offered the option of paying a bribe to be switched to the high-payment game.

The research assistant administering the session framed that opportunity as illegal to engender a moral dilemma similar to one that might arise in real life. The person mentioned that the boss was not around and that the participant could easily be switched to the high-paying game without anyone finding out. Thus, we ended up with three groups of people: low-payment no bribe, high-payment no bribe, and bribe exposed; the last group could be further split into bribe payers and bribe refusers. This arrangement allowed us to assess how ethically those exposed to the bribe would behave after having encountered the offer.

We administered three versions of the test to a total of 349 individuals in our behavioral lab. In the first two studies, some participants were offered the possibility of paying a $2 bribe to be placed in the high-payment version of the game, and 85 percent of them paid. Crucially, we observed that in the games they went on to play, bribe-exposed participants cheated more than participants who did not receive such a request. In the second study, for example, bribe-exposed participants cheated 9 percent more than those who played the high-payment version of the game and 14 percent more than participants who played the low-payment version of the game but had not been asked for a bribe.

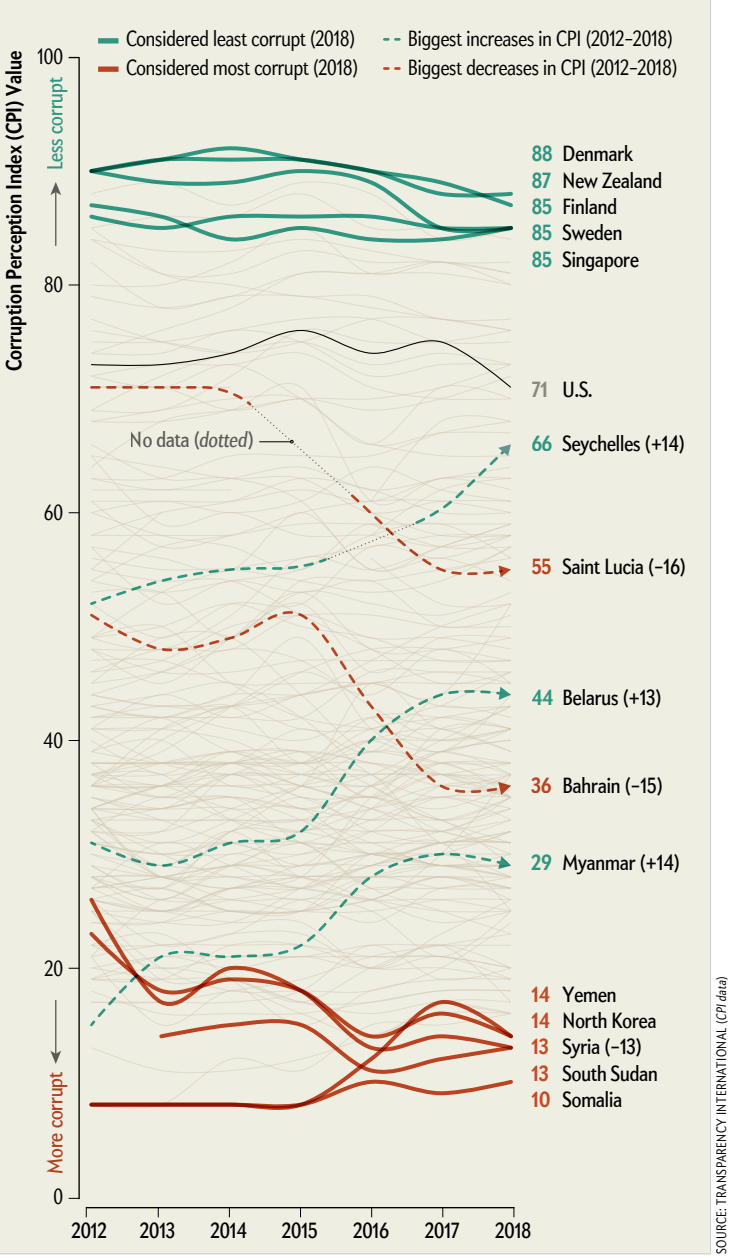
In a third study, we tested whether people act more immorally when they pay a bribe or when they are merely exposed to one. We made the bribe costlier at $12, and 82 percent turned down the request, giving us a large sample size of bribe refusers. Disturbingly, even when we limited our analysis to this group of apparently ethical individuals, we found that bribe-exposed people cheated more than those who did not receive the illegal request. Taken together, results from these three experiments suggest receiving a bribe request erodes moral character, prompting people to behave more dishonestly in subsequent ethical decisions.

ERODING NORMS

OUR WORK SUGGESTS that bribery is like a contagious disease: it spreads quickly among individuals, often by mere exposure, and becomes harder to control as time passes. This is because social norms—the patterns of behavior that are accepted as normal—impact how people will behave in many situations, including those involving ethical dilemmas. In 1991 psychologists Robert B. Cialdini, Carl A. Kallgren and Raymond R. Reno drew the important distinction between descriptive norms—the perception of what most people do—and injunctive norms—the perception of what most people approve or disapprove of. We argue that both types of norms influence bribery. Simply put, knowing that others are paying bribes to obtain preferential treatment (a descriptive norm) makes people feel that it is more acceptable to pay a bribe themselves. Similarly, thinking that others believe that paying a bribe is acceptable (an injunctive norm) will make people feel more comfortable when accepting a bribe request.

Corruption Perception Index

Levels of corruption in the public sector vary greatly around the world, according to Transparency International. Every year the nongovernmental agency uses opinion surveys and expert assessments to rank countries on a corruption scale ranging from 0 to 100. The chart displays the evolution of these rankings from 2012 to 2018, highlighting the most and least corrupt countries, as well as a few that evinced the greatest change in corruption. Levels of dishonest behavior can worsen or decline with surprising rapidity but are relatively stable in the least corrupt countries. Curiously, behavioral studies show that the innate inclination of individuals to behave dishonestly is roughly the same in different countries, regardless of their actual levels of corruption.



SOURCE: TRANSPARENCY INTERNATIONAL (*CPI data*)

Graphic by Tiffany Farrant-Gonzalez

© 2020 Scientific American

Bribery becomes normative, affecting people's moral character.

In 2009 Ariely, with behavioral researchers Francesca Gino and Shahar Ayal, published a study showing how powerful social norms can be in shaping dishonest behavior. In two lab studies, they assessed the circumstances in which exposure to others' unethical behavior would change someone's ethical decision-making. Group membership turned out to have a significant effect: When individuals observed an in-group member behaving dishonestly (a student with a T-shirt suggesting he or she was from the same school cheating in a test), they, too, behaved dishonestly. In contrast, when the person behaving dishonestly was an out-group member (a student with a T-shirt from the rival school), observers acted more honestly.

But social norms also vary from culture to culture: What is acceptable in one culture might not be acceptable in another. For example, in some societies giving gifts to clients or public officials demonstrates respect for a business relationship, whereas in other cultures it is considered bribery. Similarly, gifts for individuals in business relationships can be regarded either as lubricants of business negotiations, in the words of behavioral economists Michel André Maréchal and Christian Thöni, or as questionable business practices. And these expectations and rules about what is accepted are learned and reinforced by observation of others in the same group. Thus, in countries where individuals regularly learn that others are paying bribes to obtain preferential treatment, they determine that paying bribes is socially acceptable. Over time the line between ethical and unethical behavior becomes blurry, and dishonesty becomes the "way of doing business."

Interestingly, in cross-cultural research we published in 2016 with behavioral researchers Heather Mann, Lars Hornuf and Juan Tafurt, we found that people's underlying tendency to behave dishonestly is similar across countries. We studied 2,179 native residents in the U.S., Colombia, Portugal, Germany and China. Using a game similar to the one in our bribing studies, we observed that cheating levels in these countries were about the same. Regardless of the country, people were cheating to an extent that balanced the motive of earning money with that of maintaining a positive moral image of themselves. And contrary to commonly held beliefs (which we assessed among a different set of participants) about how these countries vary, we did not find more cheaters in countries with high corruption levels (such as Colombia) than in countries with low corruption levels (Germany).

So why do we observe huge international differences in levels of corruption and bribery? It turns out that although individuals' innate tendencies to behave honestly or otherwise are similar across countries, social norms and legal enforcement powerfully influence perceptions and behaviors. In 2007 economists Raymond Fisman and Edward Miguel published a study of parking violations among United Nations diplomats living in Manhattan. They found that diplomats from high-corruption countries accumulated more unpaid parking violations. But when enforcement authorities could confiscate diplomatic license plates of offenders, the number of unpaid violations decreased significantly. Their work suggests that cultural norms and legal enforcement are key factors in shaping ethical behavior.

PROBING DEEPER

BUT WHAT ARE the psychological mechanisms involved in the exchange of a bribe? Behavioral researchers have examined these in the lab and the field. For example, in recent research behavioral economists Uri Gneezy, Silvia Saccardo and Roel van Veldhuizen studied the psychology behind the acceptance of bribes. They conducted a lab study with 573 participants, divided into groups of three. Two participants competed for a prize by writing jokes, and the third chose the winner. The writers could bribe the referees by including $5 in an envelope when submitting their entry. Gneezy and his colleagues studied how referees reacted and how receiving a bribe distorted their judgment. They found when the referees could keep only the winner's bribe, bribes distorted their judgment, but when the referees could keep the bribe regardless of the winner, bribes no longer influenced their decision. This study suggests that people are influenced by bribes out of self-interest and not because they want to return the favor to whoever paid the bribe.

In related studies, published in 2017, Nils Köbis, now at the University of Amsterdam, and his colleagues tested the idea that severe corruption emerges gradually through a series of increasingly dishonest acts. They found that, in fact, participants in their four experiments were more likely to behave unethically when given the opportunity to do so in an abrupt manner—that is, when tempted with a single opportunity to behave unethically for a large gain rather than when faced with a series of choices for small benefits. As the researchers concluded, "sometimes the route to corruption leads over a steep cliff rather than a slippery slope."

Given how damaging corruption is to societies, we believe it is crucial to further probe its psychological roots. Three areas beg for future research. First, we need a fuller accounting of what drives a culture toward less ethical behavior. What, for example, prompts someone to ask for a bribe? What impacts the likelihood of accepting a bribe? Second, what are the consequences of bribery? Clearly, bribery and, more broadly, dishonesty are contagious. But future research could investigate the lasting effects of bribery over time and across domains: What happens when people are consistently exposed to bribes? Does recurring exposure to bribery strengthen or weaken the effect of bribes on individual dishonesty? Last, what kinds of interventions would be most effective in reducing bribe solicitations and acceptance?

Going back to our initial example, we see that the corrupt exchange that the City Hall employee offered might have seemed trivial or at least be considered an isolated event. Sadly, a single bribe request will affect the requester and the recipient. And notably, its dominolike effect can impact many individuals over time, spreading quickly across a society and, if left unchecked, entrenching a culture of dishonesty. SA

Dan Ariely is James B. Duke Professor of Psychology & Behavioral Economics at Duke University and founder of the Center for Advanced Hindsight. He is co-creator of a documentary on corruption and a bestselling author.

Ximena Garcia-Rada is a doctoral candidate of marketing at Harvard Business School. She studies how social factors influence consumer decision-making.

MORE TO EXPLORE

The (Honest) Truth about Dishonesty. Dan Ariely. HarperCollins, 2012.
(Dis)honesty: The Truth about Lies. Documentary starring Dan Ariely. Directed by Yael Melamede. Bond/360, 2015.

scientificamerican.com/magazine/sa

© 2020 Scientific American

© 2020 Scientific American

THE SCIENCE OF ANTI-SCIENCE THINKING

Convincing people who doubt the validity of climate change and evolution to change their beliefs requires overcoming a set of ingrained cognitive biases

By Douglas T. Kenrick, Adam B. Cohen, Steven L. Neuberg and Robert B. Cialdini

IN BRIEF

Scientific thinking often encounters ambivalent responses. Acclaim follows the arrival of heavier-than-air flying machines or the smartphone. But a finding challenging the political or religious status quo can trigger censure and opposition.

How do the results of evidence-based research get the hearing that they deserve? Simply presenting a litany of facts is not enough. In fact, this approach can backfire because of humans' propensity to forgo rational decision making.

Psychologists who study thinking patterns have devised strategies to counteract our tendency to take mental shortcuts, to reinforce preexisting beliefs and to succumb to the pressures exerted by other members of the various groups to which we belong.

Illustration by Heads of State

© 2020 Scientific American

IN PRINCIPLE, SCIENCE SHOULD SET ITSELF APART FROM THE HUE AND CRY of partisan bickering. After all, the scientific enterprise reaches its conclusions by testing hypotheses about the workings of the natural world. Consider the porpoise. Based on its appearance and aquatic home, the animal was assumed to be a fish. But evidence gleaned from observing its bone structure, its lack of gills and the genes it holds in common with other warm-blooded land animals leads to its being classified as a mammal with a very high level of confidence.

Yet a consensus about what constitutes a fact does not always come so readily. Take a glance at your online news feed. On a regular basis, government decision makers enact policies that fail to heed decades of evidence on climate change. In public opinion surveys, a majority of Americans choose not to accept more than a century of evidence on evolution by natural selection. Academic intellectuals put the word "science" in quotes, and members of the lay public reject vaccinations for their children.

Scientific findings have long met with ambivalent responses: A welcome mat rolls out instantly for horseless buggies or the latest smartphones. But hostility arises just as quickly when scientists' findings challenge the political or religious status quo. Some of the British clergy strongly resisted Charles Darwin's theory of evolution by natural selection. Samuel Wilberforce, bishop of Oxford, asked natural selection proponent Thomas Huxley, known as "Darwin's bulldog," on which side of his family Huxley claimed descent from an ape.

In Galileo's time, officials of the Roman Catholic Church, well-educated and progressive intellectuals in most respects, expressed outrage when the Renaissance scientist reported celestial observations that questioned the prevailing belief that Earth was the center of the universe. Galileo was placed under house arrest and forced to recant his views as heresy.

In principle, scientific thinking should lead to decisions based on consideration of all available information on a given question. When scientists encounter arguments not firmly grounded in logic and empirical evidence, they often presume that purveyors of those alternative views either are ignorant of the facts or are attempting to discourage their distribution for self-serving reasons—tobacco company executives suppressing findings linking tobacco use to lung cancer, for instance. Faced with irrational or tendentious opponents, scientists often grow increasingly strident. They respond by stating the facts more loudly and clearly in the hope that their interlocutors will make more educated decisions.

Several lines of research, however, reveal that simply presenting a litany of facts does not always lead to more objective decision making. Indeed, in some cases, this approach might actually backfire. Human beings are intelligent creatures, capable of masterful intellectual accomplishments. Unfortunately, we are not completely rational decision makers.

Understanding why people engage in irrational thinking requires combining knowledge from a range of psychological disciplines. As authors, each of us studies a separate area addressing how biased views originate. One of us (Cialdini) has expertise in heuristics, the rules that help us to quickly make everyday choices. Another of the authors (Kenrick) has studied how decisions are distorted by social motives such as the desire to find a mate or protect oneself from physical harm.

Yet another of us—Cohen—has investigated how religious beliefs affect judgment. Finally, Neuberg has studied simple cognitive biases that lead people to hold on to existing beliefs when confronted with new and conflicting evidence. All of us, in different ways, have tried to develop a deeper understanding of the psychological mechanisms that warp rationality.

Explaining why thinking goes astray is critically important to dispel false beliefs that circulate among politicians, students or even misinformed neighbors. Our own research and that of our colleagues have identified key obstacles that stand in the way of clear scientific thought. We have investigated why they arise and how they might be challenged and ultimately knocked down. Among the many hurdles, three in particular stand out:

Shortcuts. Human brains are endowed with a facile means for dealing with information overload. When we are overwhelmed or are too short on time, we rely on simple heuristics, such as accepting the group consensus or trusting an expert.

Confirmation Bias. Even with ample time and sufficient interest to move beyond shortcuts, we sometimes process information in a manner less like an impartial judge and more like a lawyer working for the mob. We show a natural tendency to pay attention to some findings over others and to reinterpret mixed evidence to fit with preexisting beliefs.

Social Goals. Even if we surmount the first two obstacles, powerful forms of social motivation can interfere with an objective analysis of the facts at hand. Whether one is biased toward reaching one scientific conclusion versus another can be influenced by the desire to win status, to conform to the views of a social network or even to attract a mate.

BEWARE THE SHORTCUT

MASTERY OF THE SCIENCES requires dealing with a set of difficult concepts. Take Darwin's theory of natural selection. To understand it, one must comprehend a set of logical premises—that limited resources favor individuals who are better able to procure food, shelter and mates, thereby leading to selective representation of traits that confer these skills to future generations.

© 2020 Scientific American

SARAH MORRIS Getty Images

MARCH FOR SCIENCE in Los Angeles, one of many held in 2017, tried to bolster support for the scientific community and for dealing with issues such as climate change. Pro-Trump counterdemonstrators also rallied.

The student of Darwinian theory must also know something about comparative anatomy (whales have bone structures more similar to humans than they do to fish). Another prerequisite is a familiarity with ecology, modern genetics and the fossil record.

Although natural selection stands out as one of the most solidly supported scientific theories ever advanced, the average citizen has not waded through textbooks full of evidence on the topic. In fact, many of those who have earned doctorates in scientific fields, even for medical research, have never taken a formal course in evolutionary biology. In the face of these challenges, most people rely on mental shortcuts or the pronouncements of experts, both strategies that can lead them astray. They may also rely—at their own peril—on intuition and gut instinct.

We use heuristics because they frequently work quite well. If a computer malfunctions, users can spend months learning about its various electronic components and how they are connected—or they can ask a computer technician. If a child develops a serious health problem, parents can study the medical literature or consult a physician.

But sometimes shortcuts serve us poorly. Consider a classic 1966 study by psychiatrist Charles K. Hofling and his colleagues on how things can go terribly wrong when people rely on the title "Dr." as a cue to an individual's authority. In the study, nurses working on a busy hospital ward received a phone call from a man who identified himself as the physician of a patient on their floor. The stranger on the phone asked the nurses on duty to go to the medicine cabinet and retrieve an unfamiliar drug called Astroten and to administer a dose twice as high as the daily maximum, violating not only the boldly stated guidelines on the label but also a hospital policy requiring handwritten prescriptions. Did the nurses balk? Ninety-five percent obeyed the unknown "doctor" without raising any questions. Indeed, they had to be stopped on their way to the patient's room with the potentially dangerous drug in hand. The nurses had unknowingly applied what is known as the authority heuristic, trusting too readily in a person in a position of responsibility.

CONFIRMATION BIAS

WHEN WE CARE ENOUGH about a topic and have the time to think about it, we move beyond simple heuristics to a more systematic analysis of the actual evidence. But even when we try hard to retain an objective perspective, our existing knowledge may still get in the way.

Abundant evidence suggests that people pay selective attention to arguments that simply reinforce their own viewpoints. They find disagreement unpleasant and are inclined to dislike the bearer of positions that run counter to their current beliefs. But what happens if intelligent individuals are forced to consider evidence on both sides of an issue?

In 1979 Charles Lord, then at Stanford University, and his colleagues conducted a study with Stanford students, who should have been able to make reasonable judgments about scientific information. The students were exposed to several rounds of scientific evidence on the deterrence of the death penalty. They might first read a description of a study that questioned whether capital punishment prevents serious crime. It compared murder rates for the year before and the year after the implementation of capital punishment in 14 states. In 11 of the states, murder rates climbed after the death penalty was established, implying that it lacks a deterrent effect.

Next, the students heard arguments from other scientists about possible weaknesses in that study's evidence. Then the orig-

© 2020 Scientific American

inal researchers came back with counterarguments. After that, the students heard about a different type of study suggesting the opposite: that capital punishment stops others from committing crimes. In it, researchers compared murder rates in 10 pairs of neighboring states with different capital punishment laws. In eight of the paired states, murder rates notched lower with capital punishment on the books, supporting the death penalty. Then students heard that evidence challenged, followed by a counterargument to that challenge.

If the students began with a strong opinion one way or the other and then performed a cold, rational analysis of the facts, they might have been expected to gravitate toward a middle ground in their views, having just heard a mix of evidence that included scientific claims that contradicted both positions for and against capital punishment. But that is not what happened. Rather students who previously favored the death penalty became even more disposed toward it, and opponents of it turned more disapproving. It became clear that students on either side of the issue had not processed the information in an evenhanded manner. Instead they believed evidence that reinforced their position was stronger, whereas refutations of that evidence were weak. So even if counterarguments can make it past our inner censors, we show an inclination to weigh those arguments in a very biased manner.

EVEN IF THE HUMAN MIND HAS MANY OBSTACLES TO OBJECTIVE THINKING, WE SHOULDN'T ACCEPT THAT IGNORANCE AND BIAS WILL ALWAYS TRIUMPH. SOCIAL PSYCHOLOGY SUGGESTS WAYS OF COPING.

A study published in September 2017, by Anthony N. Washburn and Linda J. Skitka, both then at the University of Illinois at Chicago, seems to reinforce the Stanford paper's findings. The investigators tested the hypothesis that conservatives are more distrustful of scientific evidence than liberals, perhaps because such individuals exhibit rigid thinking and are less open to new experiences. What they discovered, though, is that those on both the right and left reject scientific findings that do not jibe with their own political ideologies. The authors gave 1,347 study participants scientific evidence on six hot-button issues—climate change, gun control, health care reform, immigration, nuclear power and same-sex marriage. A cursory look at the evidence from scientific studies tended to favor one side of the issue—the absolute numbers of crimes in cities with stricter gun control might be higher than in cities without it. But a closer look at the data might give credence to the opposite view—*percentage* crime reductions in those same cities might actually be greater than they were for cities lacking gun-control laws.

If the initial hasty inspection of the data tended to favor the anti-gun-control group's expectations, members would generally look no further, content with finding results that supported their particular bias. If the results contradicted the beliefs of the gun advocates, they would scrutinize the details of the study until they discovered the numbers that suggested the opposite conclusion. If the researchers, moreover, later told one of the groups that results favored the opposite side, its members tended to be skeptical of the scientists who conducted the studies.

THE SOCIAL PRESSURE GAUNTLET

ADDITIONAL OBSTACLES ARISE from the same powerful social impulses that help us get along with others. Take the scenario of an office party where an individual's co-workers sound off erroneous claims about evolution, global warming or evidence linking vaccines to autism. Confronted with that situation, does one object or keep quiet to avoid seeming disruptive?

Research on conformity runs deep in the psychological annals. In a classic 1951 study of group dynamics, psychologist Stanley Schachter observed what happened to an individual who disagreed with the majority's consensus. After trying unsuccessfully to change the divergent opinion, other group members ended up cutting off any further communication, ostracizing the outlier. A 2003 functional magnetic resonance imaging study by Kipling D. Williams, now at Purdue University, and his colleagues found that ostracism activates the brain's dorsal anterior cingulate cortex—the same region recruited when experiencing physical pain. In a 2005 study, a team of researchers led by Gregory Berns, a neuroeconomics professor at Emory University, and his colleagues found that disagreeing with a group to which you belong is associated with increased activity in the amygdala, an area that turns on in response to different types of stress. Holding an opinion different from other group members, even a correct one, hurts emotionally. It therefore comes as no surprise that people are often reluctant to provide evidence counter to what the rest of their group believes.

Social pressures can also influence how we process new information. Group consensus may encourage us to take recourse in heuristics or to cling tightly to an opinion, all of which can interfere with objective thinking.

Our own research team conducted a study in which participants would make aesthetic judgments about a series of abstract designs and paintings and then read a passage designed to put them in either a self-protective or a romantic frame of mind. In the former condition, you might be asked to imagine being awakened by a loud sound while alone at home. As the scenario unfolded, it becomes clear that an intruder has entered the house. You imagine reaching for the phone but finding that the line is dead. A call for help receives no response. Suddenly, the door to the bedroom bursts open to reveal the dark shadow of a stranger standing there.

Alternatively, you might be randomly assigned to read an account of a romantic encounter and asked to imagine being on vacation and meeting an attractive person, then spending a romantic day with the partner that ends with a passionate kiss. Next you would enter a virtual chat room, joining three other participants to evaluate abstract images, including one you had earlier judged as of average interest. Before making the second judgment, though, you learn that this image has been rated as way below average by the other subjects.

So did study subjects change their initial judgment to conform to the other group members? How people responded depended on their current goals. Study participants who had read

© 2020 Scientific American

the home break-in scenario were more likely to conform to the group judgment. In contrast, those exposed to the amorous story answered differently depending on gender: women conformed, but men actually went against the group's judgment.

Other studies by our team have found that fear can lead both men and women to comply with group opinion, whereas sexual motives prompt men to try to stand out from the group, perhaps to show that they are worthy mates. Men, in this frame of mind, are more likely to challenge the consensus and increase the riskiness of their actions. In all cases, though, our participants' views were shaped by their social goals in the moment. They did not process available information in a completely objective way.

WHAT TO DO

IF THE HUMAN MIND is built with so many obstacles to objective scientific thinking, should we just give up and accept that ignorance and bias will always triumph? Not at all. Research in social psychology also suggests ways of coping with heuristics, confirmation biases and social pressures.

We have seen that people frequently rely on heuristics when they lack the time or interest to carefully consider the evidence. But such rules of thumb can often be defeated with simple interventions. In one experiment by market researchers Joseph W. Alba and Howard Marmorstein, subjects considered information about a dozen separate features of two cameras. Brand A was superior to brand B on just four of the features, but these were features critical in considering camera quality—the exposure accuracy, for instance. Brand B, on the other hand, came recommended as superior on eight features, all of which were relatively unimportant—having a shoulder strap, for example. Some subjects examined each attribute for only two seconds; others had more time to study all the information.

When they had only two seconds to evaluate each feature, only a few subjects (17 percent) preferred the higher-quality camera, most opting instead for the one with a greater number of unimportant functions. When the subjects were given sufficient time and allowed to directly compare the two cameras, however, more than two thirds favored the camera with the few features key to its overall quality. These results suggest that when communicating complicated evidence, sufficient time is needed to switch from a heuristic to a systematic mode of thinking that allows for better overall evaluation.

Confirmation biases can often be overcome by changing one's perspective. The same Stanford researchers who studied attitudes toward capital punishment also investigated how to change them. They instructed some students to remain objective and weigh evidence impartially in making a hypothetical decision related to the death penalty. That instruction had no effect. Others were asked to play their own devil's advocate by considering what their opinions would have been if the research about the death penalty had contradicted their own views. Biases suddenly vanished—students no longer used new evidence to bolster existing preconceptions.

One way to counteract social pressures requires first exploring whether agreement within the group really exists. Someone who disagrees with an erroneous opinion can sometimes open other group members' minds. In a 1955 *Scientific American* article, social psychologist Solomon E. Asch described studies on conformity, finding that if a single person in the group disagreed with the majority, consensus broke down. Similarly, in Stanley Milgram's famed studies of obedience—in which participants were led to believe that they were delivering painful shocks to an individual with a heart problem—blind obedience dissipated if other team members chose not to obey.

Fear increases the tendency toward conformity. If you wish to persuade others to reduce carbon emissions, take care whom you scare: a message that arouses fear of a dystopian future might work well for an audience that accepts the reality of climate change but is likely to backfire for a skeptical audience.

We have provided a few simple suggestions for overcoming psychological obstacles to objective scientific thinking. There is a large literature on persuasion and social influence that could be quite useful to anyone attempting to communicate with a group holding beliefs that fly in the face of scientific evidence. For their part, scientists need to adopt a more systematic approach in collecting their own data on the effectiveness of different strategies for confronting antiscientific thinking about particular issues. It is essential to understand whether an individual's resistance to solid evidence is based on simple heuristic thinking, systematic bias or particular social motives.

These steps are critical because antiscientific beliefs can lead to reduced research funding and a consequent failure to fully understand potentially important phenomena that affect public welfare. In recent decades government funding has decreased for research into the health impact of keeping guns in the home and of reducing the harmful effects of air pollution. Guns in the home are frequently involved in teenage suicides, and an overwhelming scientific consensus shows that immediate measures are needed to address the planet's warming.

It is easy to feel helpless in the face of our reluctance to embrace novel scientific findings. Still, there is room for optimism: the majority of Galileo's fellow Italians and even the pope now accept that our planet revolves around the sun, and most of Darwin's compatriots today endorse the theory of evolution. Indeed, the Anglican Church's director of public affairs wrote an apology to Darwin for the 200th anniversary of his birth. If scientists can incorporate the insights of research on the psychological obstacles to objective thinking, more people will accept objective evidence of how the natural world functions as well. SA

Douglas T. Kenrick is a professor of psychology at Arizona State University, who has studied behaviors ranging from altruism to homicidal fantasies.

Adam B. Cohen is a professor of psychology at Arizona State, whose work focuses on the psychological foundations of religious beliefs.

Steven L. Neuberg is a Foundation Professor and chair of the department of psychology at Arizona State. He researches stereotyping, prejudice and the effects of religion on conflict.

Robert B. Cialdini is Regents' Professor Emeritus of psychology and marketing at Arizona State. He explores the reasons that people comply with requests in everyday settings.

MORE TO EXPLORE

Influence: The Psychology of Persuasion. Revised edition. Robert B. Cialdini. Harper Business, 2009.
The Rational Animal: How Evolution Made Us Smarter Than We Think. Douglas T. Kenrick and Vladas Griskevicius. Basic Books, 2013.
Mindware: Tools for Smart Thinking. Richard E. Nisbett. Farrar, Straus and Giroux, 2015.
Attitude Roots and Jiu Jitsu Persuasion: Understanding and Overcoming the Motivated Rejection of Science. Matthew J. Hornsey and Kelly S. Fielding in *American Psychologist,* Vol. 72, No. 5, pages 459–473; July–August 2017.

scientificamerican.com/magazine/sa

© 2020 Scientific American

How Professional Truth Seekers Search for Answers

Nine experts describe how they sort signal from noise

As told to Brooke Borel

Illustrations by Bud Cook

HOW A DATA JOURNALIST SEARCHES FOR ANSWERS

People assume that because there are data, that the data must be true. But the truth is, all data are dirty.

People create data, which means data have flaws just like people. One thing data journalists do is interrogate that assumption of truth, which serves an important accountability function—a power check to make sure we aren't collectively getting carried away with data and making bad social decisions.

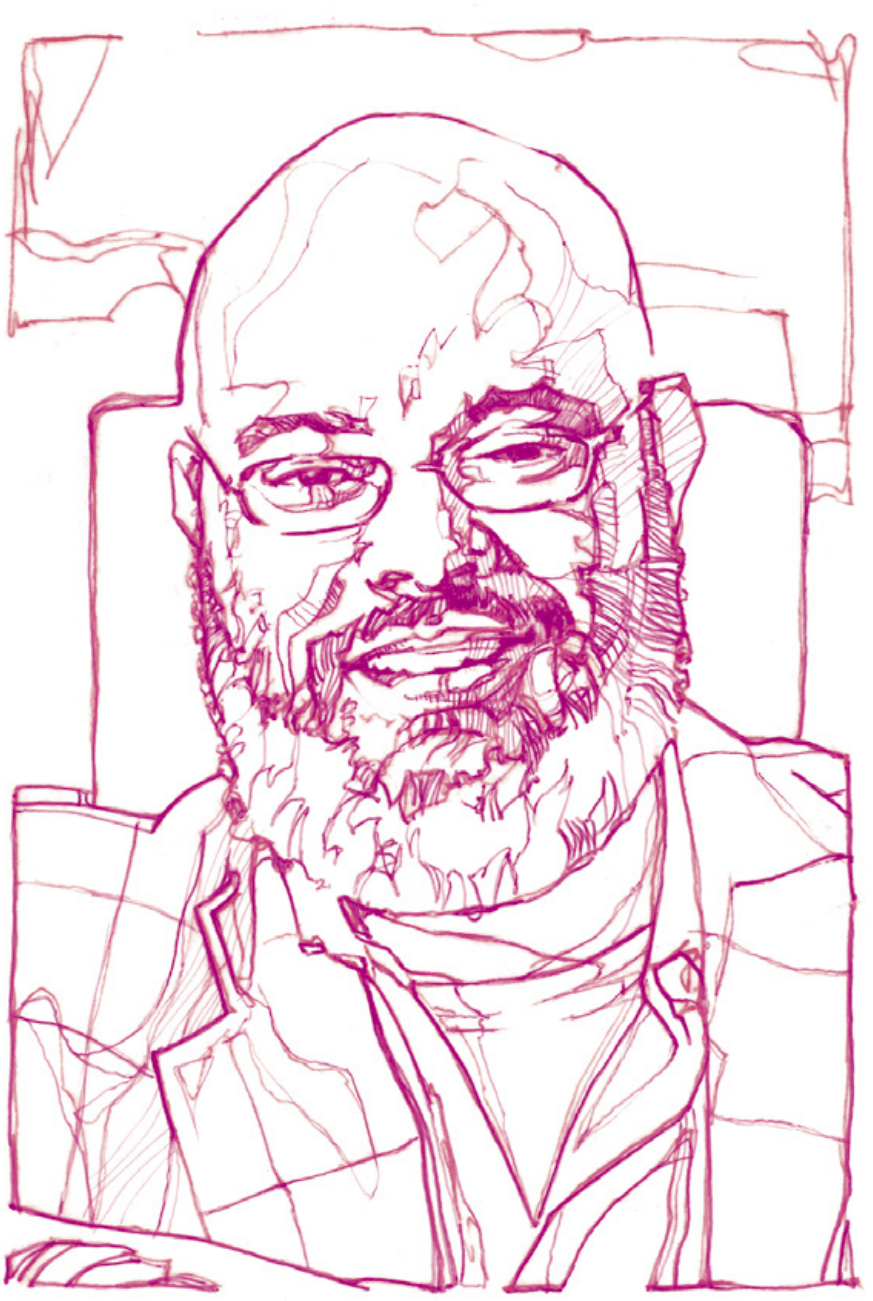
To interrogate the data, you have to do a lot of janitorial work. You have to clean and organize them; you have to check the math. And you also have to acknowledge the uncertainty. If you are a scientist, and you don't have the data, you can't write the paper. But one of the fabulous things about being a data journalist is that sparse data don't deter us—sometimes the lack of data tells me something just as interesting. As a journalist, I can use words, which are a magnificent tool for communicating about uncertainty.

Meredith Broussard, an associate professor at the Arthur L. Carter Journalism Institute at New York University

© 2020 Scientific American

HOW A BEHAVIORAL SCIENTIST SEARCHES FOR ANSWERS

The kind of control you have in bench science is much tighter than in behavioral science—



the power to detect small effects in people is much lower than in, say, chemistry. Not only that, people's behaviors change across time and culture. When we think about truth in behavioral science, it's really important not only to reproduce a study directly but also to extend reproduction to a larger number of situations—field studies, correlational studies, longitudinal studies.

So how do we measure racism, something that's not a single behavior but a pattern of outcomes—a whole system by which people are oppressed? The best approach is to observe the pattern of behaviors and then see what happens when we alter or control for a variable. How does the pattern change? Take policing. If we remove prejudice from the equation, racially disparate patterns persist. The same is true of poverty, education and a host of things we think predict crime. None of them are sufficient to explain patterns of racially disparate policing outcomes. That means we still have work to do. Because it's not like we don't know how to produce nonviolent and equitable policing. Just look at the suburbs. We've been doing it there for generations.

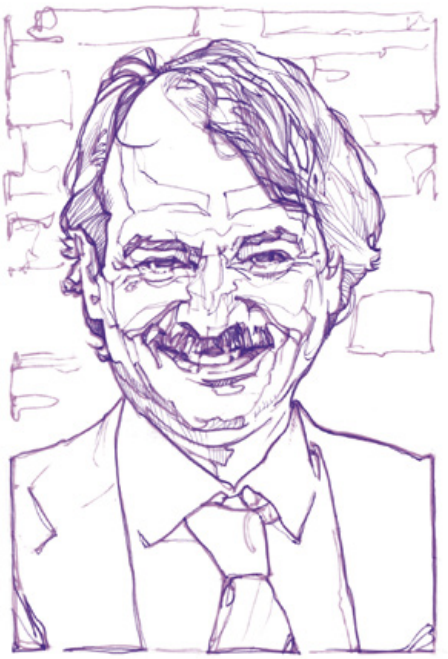
Of course, there is uncertainty. In most of this world, we are nowhere near confidence about causality. Our responsibility as scientists is to characterize these uncertainties because a wrong calculation in what drives something like racism is the difference between getting policies right and getting them wrong.

Phillip Atiba Goff, *Franklin A. Thomas Professor in Policing Equity at the John Jay College of Criminal Justice at the City University of New York and president of the Center for Policing Equity*

© 2020 Scientific American

HOW A PHYSICIAN SEARCHES FOR ANSWERS

The answer to questions about human life isn't a certain thing,



like measuring how a stone drops to the ground in exactly so many seconds. If it were, it probably would not be life. It would be a stone. Within biomedicine, it's tricky finding out if an effect is real—there are different standards across different fields. Not all tools will work for every question, and there are different levels of complexity for what we know before we even start a study.

Still, the one core dimension across biomedicine is the ability to replicate, in a new study, what was seen in the first investigation. For many years in the field, we have been discouraged from doing this. Why waste money to do the exact thing you had done before, let alone something someone else had done before? But many researchers are realizing it is not possible to leave out replication studies.

To make replication work, though, it is essential to have a detailed explanation of how the original study was done. You need the instructions, the raw data and maybe even some custom-built computer software. For a long time, scientists didn't want to share that information, but that is changing. Science is a communal effort, and we should default to being open and sharing.

John P. A. Ioannidis, a professor of medicine at Stanford University

HOW A SOCIAL TECHNOLOGIST SEARCHES FOR ANSWERS

The biggest epistemological question facing the field of machine learning is: What is our ability to test a hypothesis?

Algorithms learn to detect patterns and details from massive sets of examples—for instance, an algorithm could learn to identify a cat after seeing thousands of cat photographs. Until we have greater interpretability, we can test how a result was achieved by appealing conclusions from the algorithms. This raises the specter that we don't have real accountability for the results of deep-learning systems—let alone due process when it comes to their effects on social institutions. These issues are part of a live debate in the field.

Also, does machine learning represent a type of rejection of the scientific method, which aims to find not only correlation but also causation? In many machine-learning studies, correlation has become the new article of faith, at the cost of causation. That raises real questions about verifiability.

In some cases, we may be taking a step backward. We see this in the space of machine vision and affect recognition. These are systems that extrapolate from photographs of people to predict their race, gender, sexuality or likelihood of being a criminal. These sorts of approaches are both scientifically and ethically concerning—with echoes of phrenology and physiognomy. The focus on correlation should raise deep suspicions in terms of our ability to make claims about people's identity. That's a strong statement, by the way, but given the decades of research on these issues in the humanities and social sciences, it should not be controversial.

Kate Crawford, a distinguished research professor at New York University, co-founder of the AI Now Institute at N.Y.U. and member of Scientific American*'s board of advisers*

© 2020 Scientific American

HOW A STATISTICIAN SEARCHES FOR ANSWERS

In statistics, we aren't generally seeing the whole universe but only a slice of it.

A small slice usually, which could tell a completely different story than another small slice. We are trying to make a leap from these small slices to a bigger truth. A lot of people take that basic unit of truth to be the *p*-value, a statistical measure of how surprising what we see in our small slice is, if our assumptions about the larger universe hold. But I don't think that's correct.

In reality, the notion of statistical significance is based on an arbitrary threshold applied to the *p*-value, and it may have very little to do with substantive or scientific significance. It's too easy to slip into a thought pattern that provides that arbitrary threshold with meaning—it gives us a false sense of certainty. And it's also too easy to hide a multitude of scientific sins behind that *p*-value.

One way to strengthen the *p*-value would be to shift the culture toward transparency. If we not only report the *p*-value but also show the work on how we got there—the standard error, the standard deviation or other measures of uncertainty, for example—we can give a better sense of what that number means. The more information we publish, the harder it is to hide behind that *p*-value. Whether we can get there, I don't know. But I think we should try.

Nicole Lazar, *a professor of statistics at the University of Georgia*

HOW A NEUROSCIENTIST SEARCHES FOR ANSWERS

Science does not search for truth, as many might think.

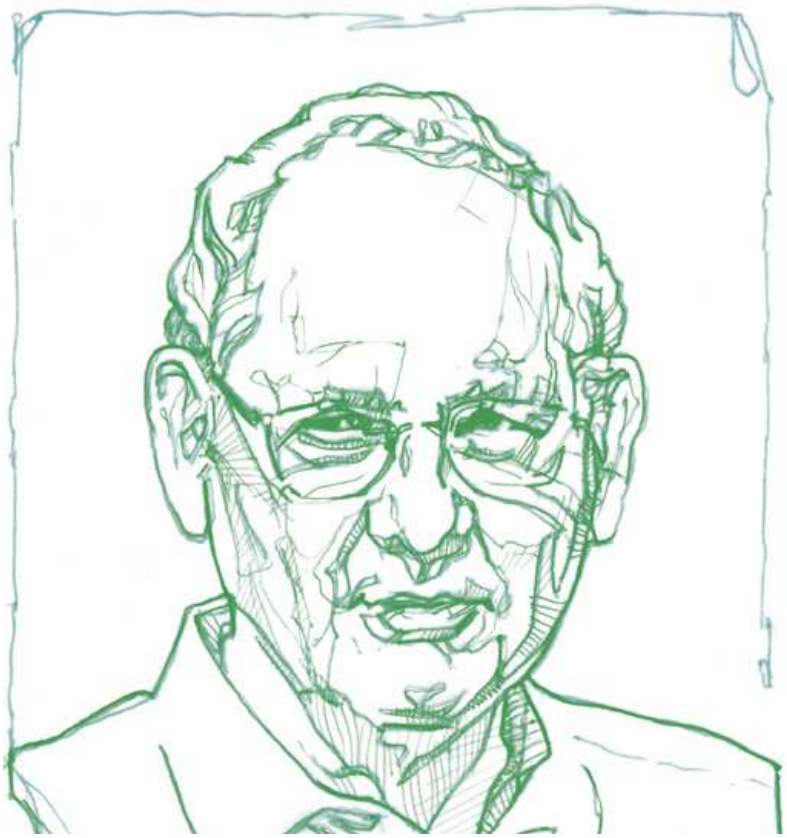
Rather the real purpose of science is to look for better questions. We run experiments because we are ignorant about something and want to learn more, and sometimes those experiments fail. But what we learn from our ignorance and failure opens new questions and new uncertainties. And these are better questions and better uncertainties, which lead to new experiments. And so on.

Take my field, neurobiology. For around 50 years the fundamental question for the sensory system has been: What information is being sent into the brain? For instance, what do our eyes tell our brain? Now we are seeing a reversal of that idea: the brain is actually asking questions of the sensory system. The brain may not be simply sifting through massive amounts of visual information from, say, the eye; instead it is asking the eye to seek specific information.

In science, there are invariably loose ends and little blind alleys. While you may think you have everything cleared up, there is always something new and unexpected. But there is value in uncertainty. It shouldn't create anxiety. It's an opportunity.

Stuart Firestein, *a professor in the department of biological sciences at Columbia University*

© 2020 Scientific American



HOW A HISTORICAL LINGUIST SEARCHES FOR ANSWERS

Like any scientist, linguists rely on the scientific method.

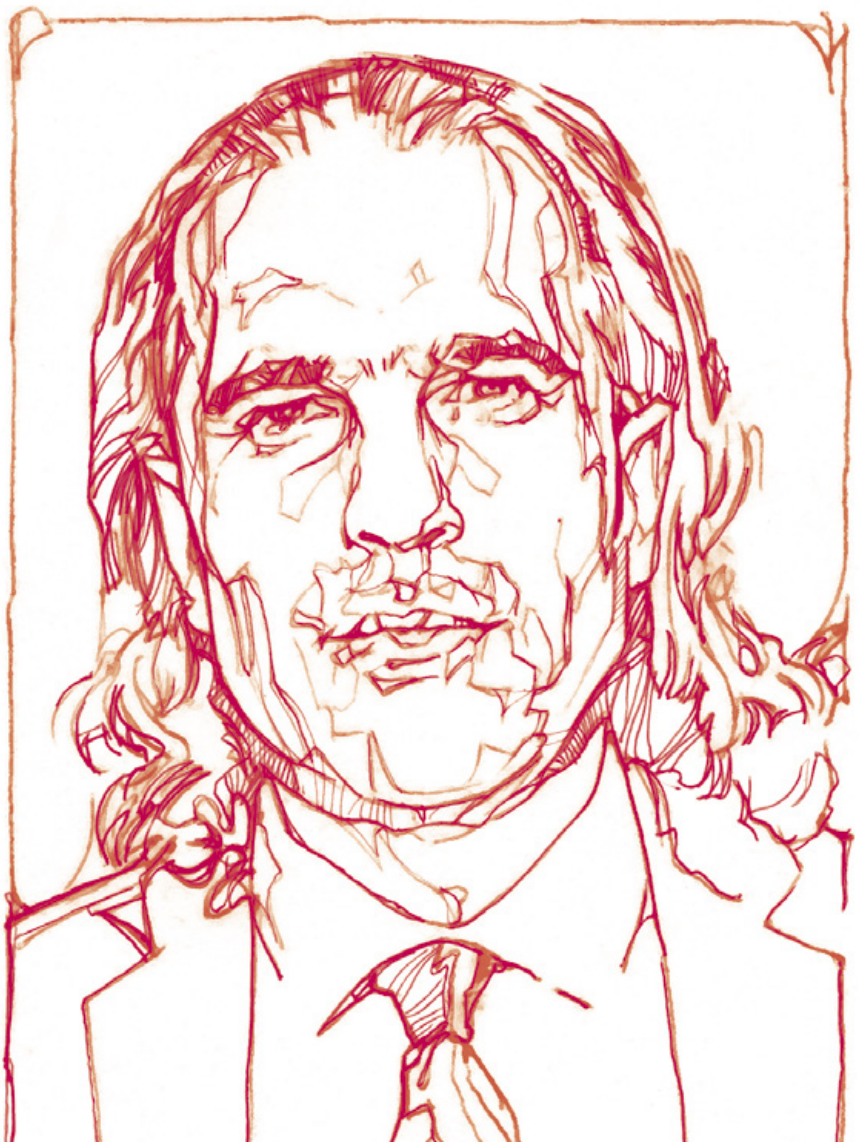
One of the principal goals of linguistics is to describe and analyze languages to discover the full range of what is possible and not possible in human languages. From this, linguists aim to reach their goal of understanding human cognition through the capacity for human language.

So there is an urgency to efforts to describe endangered languages, to document them while they are still in use, to determine the full range of what is linguistically possible. There are around 6,500 known human languages; around 45 percent of them are endangered.

Linguists use a specific set of criteria to identify endangered languages and to determine just how

HOW A THEORETICAL PHYSICIST SEARCHES FOR ANSWERS

Physics is the most mature science, and physicists are obsessive on the subject of truth.



There is an actual universe out there. The central miracle is that there are simple underlying laws, expressed in the precise language of mathematics, which can describe it. That said, physicists don't traffic in certainties but in degrees of confidence. We've learned our lesson: throughout history, we have again and again found out that some principle we thought was central to the ultimate description of reality isn't quite right.

To figure out how the world works, we have theories and build experiments to test them. Historically, this method works. For example, physicists predicted the existence of the Higgs boson particle in 1964, built the Large Hadron Collider

© 2020 Scientific American

endangered a language is: Are children still learning the language? How many individual people speak it? Is the percentage of speakers declining with respect to the broader population? And are the contexts in which the language is used decreasing?

The question of scientific objectivity and "truth" is connected to endangered language research. Truth, in a way, is contextual. That is, what we hold to be true can change as we get more data and evidence or as our methods improve. The investigation of endangered languages often discovers things that we did not know were possible in languages, forcing us to reexamine previous claims about the limits of human language, so that sometimes what we thought was true can shift.

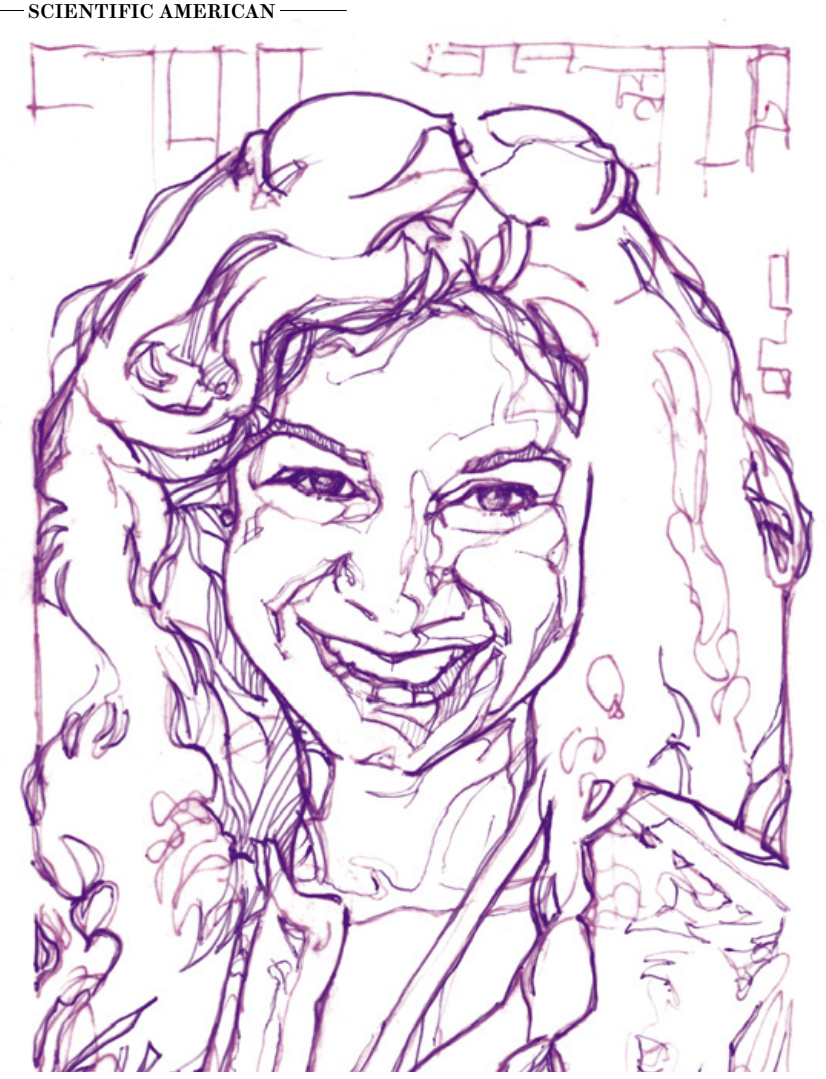
Lyle Campbell, an emeritus professor of linguistics at the University of Hawaii at Mānoa

(LHC) at CERN in the late 1990s and early 2000s, and found physical evidence of the Higgs in 2012. Other times we can't build the experiment—it is too massive or expensive or would be impossible with available technology. So we try thought experiments that pull from the existing infrastructure of existing mathematical laws and experimental data.

Here's one: The concept of spacetime has been accepted since the early 1900s. But to look at smaller spaces, you have to use more powerful resolution. That's why the LHC is 17 miles around—to produce the huge energies needed to probe tiny distances between particles. But at some point, something bad happens. You'll put out such an enormous amount of energy to look at such a small bit of space that you'll actually create a black hole instead. Your attempt to see what is inside makes it impossible to do so, and the notion of spacetime breaks down.

At any moment in history, we can understand some aspects of the world but not everything. When a revolutionary change brings in more of the larger picture, we have to reconfigure what we knew. The old things are still part of the truth but have to be spun around and put back into the larger picture in a new way.

Nima Arkani-Hamed, a professor in the School of Natural Sciences at the Institute for Advanced Study in Princeton, N.J.



HOW A PALEOBIOLOGIST SEARCHES FOR ANSWERS

Our basic unit of truth in paleobiology is the fossil—

a clear record of life in the past—and we also use genetic evidence from living organisms to help us put fossils within the tree of life. Together they help us understand how these creatures changed and how they are related. Because we are looking at extinct animals as they existed in a broader ecosystem, we pull in information from other fields: chemical analysis of surrounding rocks to get a sense of the fossil's age, where the world's landmasses might have been at the time, what kind of environmental changes were happening, and so on.

To discover fossils, we scour the landscape to find them among rocks. You can tell the difference between a fossil and any old rock by its shape and its internal structure. For example, a fossil bone will have tiny cylinders called osteons where blood vessels once ran through the bone. Some fossils are obvious: a leg of a dinosaur, a giant, complete bone. Smaller bits can be telling, too. For mammals, which I study, you can tell a lot from the shape of a single tooth. And we can combine this information with genetics, by using DNA samples from living creatures that we think are related to the fossils, based on anatomy and other clues.

We just don't do these investigations to reconstruct past worlds but also to see what they can tell us about the world we live in. There was a huge spike in temperature 55 million years ago, for example. It was nothing like today, but still, we've found radical changes in the animals and plants from that era. We can compare those changes to see how related creatures may respond to current climate change.

Anjali Goswami, a professor and research leader at the Natural History Museum in London

© 2020 Scientific American

The Truth about Scientific Models

They don't necessarily try to predict what will happen—but they can help us understand possible futures

By Sabine Hossenfelder

As COVID-19 claims more victims, scientific models make headlines. We need these models to make informed decisions. But how can we tell whether a model can be trusted? The philosophy of science, it seems, has become a matter of life or death. Whether we are talking about traffic noise from a new highway or about climate change or a pandemic, scientists rely on models, which are simplified, mathematical representations of the real world. Models are approximations and omit details, but a good model will robustly output the quantities it was developed for.

Models do not always predict the future. This does not make them unscientific, but it makes them a target for science skeptics. I cannot even blame the skeptics, because scientists frequently praise correct predictions to prove a model's worth. It isn't originally their idea. Many eminent philosophers of science, including Karl Popper and Imre Lakatos, opined that correct predictions are a way of telling science from pseudoscience.

But correct predictions alone don't make for a good scientific model. And the opposite is also true: a model can be good science without ever making predictions. Indeed, the models that matter most for political discourse are those that do *not* make predictions. Instead they produce "projections" or "scenarios" that, in contrast to predictions, are forecasts that depend on the course of action we will take. That is, after all, the reason we consult models: so we can decide what to do. But because we cannot predict political decisions themselves, the actual future trend is necessarily unpredictable.

This has become one of the major difficulties in explaining pandemic models. Dire predictions in March for COVID's global death toll have not come true. But they were projections for the case in which we take no measures; they were not predictions.

Political decisions are not the only reason why a model may merely make contingent projections rather than definite predictions. Trends of global warming, for example, depend on the frequency and severity of volcanic eruptions, which themselves cannot currently be predicted. They also depend on technological progress, which itself depends on economic prosperity, which again depends, among many other things, on whether society is in the grasp of a pandemic. Sometimes asking for predictions is really asking for too much.

Predictions are also not enough to make for good science. Recall how each time a natural catastrophe happens, it turns out to have been "predicted" in a movie or a book. Given that most natural catastrophes are predictable to the extent that "eventually something like this will happen," this is hardly surprising. But these are not predictions, they are scientifically meaningless prophecies because they are not based on a model whose methodology can be reproduced, and no one has tested whether the prophecies were better than random guesses.

Thus, predictions are neither necessary for a good scientific model nor sufficient to judge one. But why, then, were the philosophers so adamant that good science needs to make predictions? It's not that they were wrong. It's just that they were trying to address a different problem than what we are facing now.

Scientists tell good models from bad ones by statistical methods that are hard to communicate without equations. These methods depend on the type of model, the amount of data and the field of research. In short, it's difficult. The rough answer is that a good scientific model accurately explains a lot of data with few assumptions. The fewer the assumptions and the better the fit to data, the better the model.

But the philosophers were not concerned with quantifying explanatory power. They were looking for a way to tell good science from bad science without

© 2020 Scientific American

GETTY IMAGES

having to dissect scientific details. And while correct predictions may not tell you whether a model is good science, they increase trust in the scientists' conclusions because predictions prevent scientists from adding assumptions after they have seen the data. Thus, asking for predictions is a good rule of thumb, but it is a crude and error-prone criterion. And fundamentally it makes no sense. A model either accurately describes nature, or it doesn't. At which moment in time a scientist made a calculation is irrelevant for the model's relation to nature.

A confusion closely related to the idea that good science must make predictions is the belief that scientists should not update a model when new data come in. This can also be traced back to Popper & Co., who thought it is bad scientific practice. But of course, a good scientist updates their model when they get new data! This is the essence of the scientific method: When you learn something new, revise. In practice, this usually means recalibrating model parameters with new data. This is why we see regular updates of COVID case projections. What a scientist is not supposed to do is add so many assumptions that their model can fit any data. This would be a model with no explanatory power.

Understanding the role of predictions in science also matters for climate models. These models have correctly predicted many observed trends, from the increase of surface temperature, to stratospheric cooling, to sea ice melting. This fact is often used by scientists against climate change deniers. But the deniers then come back with some papers that made wrong predictions. In response, the scientists point out the wrong predictions were few and far between. The deniers counter there may have been all kinds of reasons for the skewed number of papers that have nothing to do with scientific merit. Now we are counting heads and quibbling about the ethics of scientific publishing rather than talking science. What went wrong? Predictions are the wrong argument.

A better answer to deniers is that climate models explain loads of data with few assumptions. The computationally simplest explanation for our observations is that the trends are caused by human carbon dioxide emission. It's the hypothesis that has the most explanatory power.

In summary, to judge a scientific model, do not ask for predictions. Ask instead to what degree the data are explained by the model and how many assumptions were necessary for this. And most of all, do not judge a model by whether you like what it tells you. SA

Sabine Hossenfelder is a physicist and research fellow at the Frankfurt Institute for Advanced Studies in Germany. She currently works on dark matter and the foundations of quantum mechanics.

MORE TO EXPLORE

Evaluating the Performance of Past Climate Model Projections. Zeke Hausfather, Henri F. Drake, Tristan Abbott and Gavin A. Schmidt in *Geophysical Research Letters*, Vol. 47, No. 1, Article e2019GL085378; January 16, 2020.

scientificamerican.com/magazine/sa

© 2020 Scientific American

© 2020 Scientific American

A Significant Problem

Standard scientific methods are under fire. Will anything change?

By Lydia Denworth

IN 1925 BRITISH GENETICIST AND STATISTICIAN RONALD Fisher published a book called *Statistical Methods for Research Workers*. The title doesn't scream "best seller," but the book was a huge success and established Fisher as the father of modern statistics. In it, he tackles the problem of how researchers can apply statistical tests to numerical data to draw conclusions about what they have found and determine whether it is worth pursuing. He references a statistical test that summarizes the compatibility of data with a proposed model and produces a p value. Fisher suggests that researchers might consider a p value of 0.05 as a handy guide: "It is convenient to take this point as a limit in judging whether a deviation ought to be considered significant or not." Pursue results with p values below that threshold, he advises, and do not spend time on results that fall above it. Thus was born the idea that a value of p less than 0.05 equates to what is known as statistical significance—a mathematical definition of "significant" results.

IN BRIEF

The use of *p* values for nearly a century to determine statistical significance of experimental results has contributed to an illusion of certainty and reproducibility crises in many scientific fields.

There is growing determination to reform statistical analysis, but researchers disagree on whether it should be tweaked or overhauled. Some suggest changing statistical methods, whereas others would do away with a threshold for defining "significant" results.

Ultimately the *p* value plays into the human need for certainty. So it may be time for both scientists and the public to embrace the discomfort of being unsure.

Illustration by Mark Allen Miller

© 2020 Scientific American

Nearly a century later, in many fields of scientific inquiry, a p value less than 0.05 is considered the gold standard for determining the merit of an experiment. It opens the doors to the essentials of academia—funding and publication—and therefore underpins most published scientific conclusions. Yet even Fisher understood that the concept of statistical significance and the p value that underpins it has considerable limitations. Most have been recognized for decades. "The excessive reliance on significance testing," wrote psychologist Paul Meehl in 1978, "[is] a poor way of doing science." *P* values are regularly misinterpreted, and statistical significance is not the same thing as practical significance. Moreover, the methodological decisions required in any study make it possible for an experimenter, consciously or unconsciously, to shift a p value up or down. "As is often said, you can prove anything with statistics," says statistician and epidemiologist Sander Greenland, professor emeritus at the University of California, Los Angeles, and one of the leading voices for reform. Studies that rely only on achieving statistical significance or pointing out its absence regularly result in inaccurate claims—they show things to be true that are false and things to be false that are true. After Fisher had retired to Australia, he was asked whether there was anything in his long career he regretted. He is said to have snapped, "Ever mentioning 0.05."

In the past decade the debate over statistical significance has flared up with unusual intensity. One publication called the flimsy foundation of statistical analysis "science's dirtiest secret." Another cited "numerous deep flaws" in significance testing. Experimental economics, biomedical research and especially psychology have been engulfed in a controversial replication crisis, in which it has been revealed that a substantial percentage of published findings are not reproducible. One of the more notorious examples is the idea of the power pose, the claim that assertive body language changes not just your attitude but your hormones, which was based on one paper that has since been repudiated by one of its authors. A paper on the economics of climate change (by a skeptic) "ended up having almost as many error corrections as data points—no kidding!—but none of these error corrections were enough for him to change his conclusion," wrote statistician Andrew Gelman of Columbia University on his blog, where he regularly takes researchers to task for shoddy work and an unwillingness to admit the problems in their studies. "Hey, it's fine to do purely theoretical work, but then no need to distract us with data," Gelman wrote.

The concept of statistical significance, though not the only factor, has emerged as an obvious part of the problem. In the past four years, hundreds of researchers have urgently called for reform, authoring or endorsing papers in prestigious journals on redefining statistical significance or abandoning it altogether. The American Statistical Association (ASA), which put out a strong and unusual statement on the issue in 2016, argues for "moving to a world beyond $p < 0.05$." Ronald Wasserstein, the ASA's executive director, puts it this way: "Statistical significance is supposed to be like a right swipe on Tinder. It indicates just a certain level of interest. But unfortunately, that's not what statistical significance has become. People say, 'I've got 0.05, I'm good.' The science stops."

The question is whether anything will change. "Nothing is new. That needs to sober us about the prospect that maybe this time will be the same as every other time," says behavioral economist Daniel Benjamin of the University of Southern California, another voice for reform. Still, although they disagree over the remedies, it is striking how many researchers do agree, as econ-

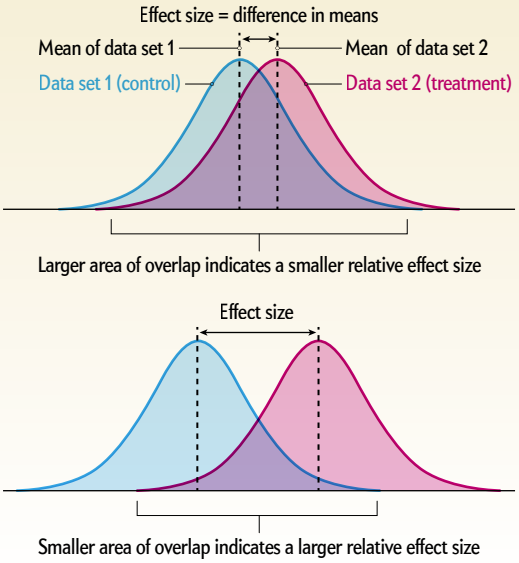
Statistical Significance

Imagine you grow pumpkins in your garden. Would using fertilizer affect their size? Given your long experience without fertilizer, you know how much the weights of pumpkins vary and you know that their average weight is 10 pounds. You decide to grow a sample of 25 pumpkins with fertilizer. The average weight of these 25 pumpkins turns out to be 13.2 pounds. How do you decide whether the difference of 3.2 pounds from the status quo of 10 pounds—the hypothetical "null" value—happened by chance or that fertilizer does indeed grow larger pumpkins?

Statistician Ronald Fisher's solution to this puzzle involves performing a thought experiment: imagine that you were to repeatedly grow 25 pumpkins a very large number of times. Each time you would get a different average weight because of the random variability of individual pumpkins. Then you would plot the distribution of those averages and consider the probability (**p value**) that the data you have generated would be possible if the fertilizer had no effect. By convention, a p value of 0.05 became a cut-off to identify significant results—in this case, ones that lead a researcher to conclude the fertilizer does not have an effect. Here we break down some of the concepts that drive the thought experiment for statistical significance.

EFFECT SIZE

The effect size for a treatment is the difference between the average outcome when the treatment is used compared with the average when the treatment is not used. The concept can be used to compare averages in samples or "true" averages for entire distributions. The effect size can be measured in the same units (such as pounds of pumpkins) as the outcome. But for many outcomes—such as responses to some psychological questionnaires—there is not a natural unit. In that case, researchers can use relative effect sizes. One way of measuring relative effect size is based on the overlap between the control and the treatment distributions.


Effect size = difference in means
Mean of data set 1
Mean of data set 2
Data set 1 (control)
Data set 2 (treatment)
Larger area of overlap indicates a smaller relative effect size
Effect size
Smaller area of overlap indicates a larger relative effect size

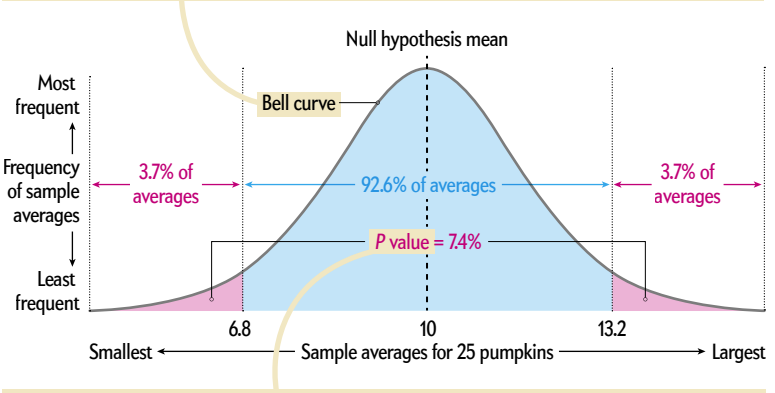
Graphic by Amanda Montañez (graphs) and Heather Krause

© 2020 Scientific American

P VALUE

To calculate the *p* value, we need to compare the actual average of 13.2 pounds that we observed in our sample of 25 pumpkins with the random distribution of averages if we were to take many new samples of 25 pumpkins.

The bell curve shows the distribution of random average weights for samples of 25 under the null hypothesis that the fertilizer has no effect.

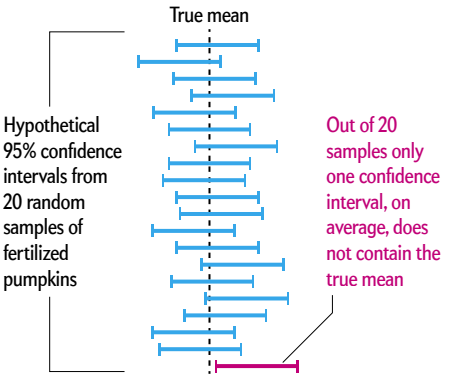


The *p* value is the probability of getting a random average weight as far from 10 as the average you actually observed, 13.2. Since 13.2 – 10 = 3.2, we want the probability of getting an average ≥ 13.2 or ≤ 6.8 (6.8 = 10 – 3.2). In this example, that probability is 0.074, which is the actual observed *p* value for your sample. Because it is greater than 0.05, your result would not be considered significant evidence that the fertilizer makes a difference.

The example shows a "two-tailed test," where the *p* value counts the probability of a weight greater than 13.2 and that of a weight less than 6.8 (10 – 3.2 = 6.8). Under some circumstances, a researcher might choose to perform a "one-tailed test." In that case, the *p* value would be only 0.037, which, being less than 0.05, is considered significant. This illustrates one way in which researchers can modify their stated intention for a study to achieve different *p* values with exactly the same data.

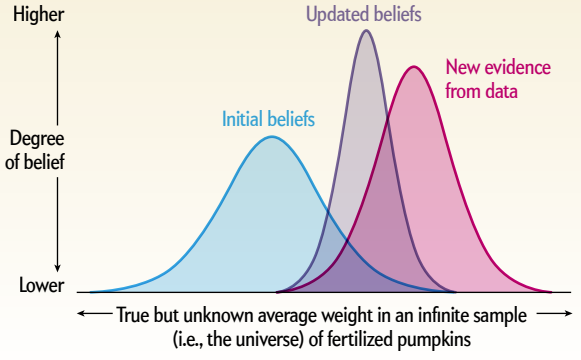
CONFIDENCE INTERVAL

We can calculate a 95 percent confidence interval from our sample of 25 pumpkins. This is a guess for the average weight of fertilized pumpkins. Calculating the 95 percent confidence interval involves inverting the calculation for the *p* value to find all hypothetical values that produce a *p* value ≥ 0.05. With our sample of 25 pumpkins, our 95 percent confidence interval goes from 9.69 to 16.71. The "true" average weight of fertilized pumpkins may or may not be in that interval. We can't be sure, so what does the "95 percent" mean? Imagine what would happen if we repeatedly grew batches of 25 pumpkins and sampled them. Each sample would produce a randomly different confidence interval. We know that in the long run, 95 percent of these intervals would include the true value and 5 percent would not. But what about our particular interval from the first pumpkin sample? We don't know whether it is in the 95 percent that worked or in the 5 percent that missed. It is the *process* that is right 95 percent of the time.


True mean
Hypothetical 95% confidence intervals from 20 random samples of fertilized pumpkins
Out of 20 samples only one confidence interval, on average, does not contain the true mean

BAYESIAN METHODS

In the Bayesian approach to inference, a person's state of uncertainty about an unknown quantity is represented by a probability distribution. Bayes' theorem is used to combine individuals' initial beliefs—their distribution before looking at data—with the information they receive from the data, which produces a mathematically implied distribution for their updated beliefs. The updated beliefs from one study become the new initial beliefs for the next study, and so on. A major area of discussion and controversy concerns attempts to find "objective" criteria for initial beliefs. The goal is to find ways of constructing initial beliefs, known as prior distributions, that can be widely accepted by researchers as reasonable.


Higher
Degree of belief
Lower
Updated beliefs
New evidence from data
Initial beliefs
True but unknown average weight in an infinite sample (i.e., the universe) of fertilized pumpkins

SURPRISAL

The *p* value conveys how surprising our pumpkin data are if we suppose that, in reality, fertilizing has no effect on growth. Some researchers have suggested that the *p* values do not convey surprisingness in a way that is intuitive for most people. Instead they suggest a mathematical quantity called a surprisal, also known as an *s* value or Shannon transform, that adjusts *p* values to produce bits (as in computer bits). Surprisal can be interpreted through the example of tossing coins.

Two heads in a row = **2** bits of surprisal = *p* value of $1/2^2$ = **0.25**

Four heads in a row = **4** bits of surprisal = *p* value of $1/2^4$ = **0.0625**

Five heads in a row = **5** bits of surprisal = *p* value of $1/2^5$ = **0.03215**

Our sample of 25 pumpkins with an average weight of 13.2 and a *p* value of 0.074 produces between 3 and 4 bits of surprisal. To be exact: 3.76 bits of surprisal since $3.76 = -\log_2(0.074)$.

© 2020 Scientific American

omist Stephen Ziliak wrote, that "the current culture of statistical significance testing, interpretation, and reporting has to go."

THE WORLD AS IT IS

THE GOAL OF SCIENCE is to describe what is true in nature. Scientists use statistical models to infer that truth—to determine, for instance, whether one treatment is more effective than another or whether one group differs from another. Every statistical model relies on a set of assumptions about how data are collected and analyzed and how the researchers choose to present their results.

Those results nearly always center on a statistical approach called null hypothesis significance testing, which produces a *p* value. This testing does not address the truth head-on; it glances at it obliquely. That is because significance testing is intended to indicate only whether a line of research is worth pursuing further. "What we want to know when we run an experiment is how likely is it [our] hypothesis is true," Benjamin says. "But [significance testing] answers a convoluted alternative question, which is, if my hypothesis were false, how unlikely would my data be?"

Sometimes this works. The search for the Higgs boson, a particle first theorized by physicists in the 1960s, is an extreme but useful example. The null hypothesis was that the Higgs boson did not exist; the alternative hypothesis was that it must exist. Teams of physicists at CERN's Large Hadron Collider ran multiple experiments and got the equivalent of a *p* value so vanishingly small that it meant the possibility of their results occurring if the Higgs boson did not exist was one in 3.5 million. That made the null hypothesis untenable. Then they double-checked to be sure the result wasn't caused by an error. "The only way you could be assured of the scientific importance of this result, and the Nobel Prize, was to have reported that [they] went through hoops of fire to make sure [none] of the potential problems could have produced such a tiny value," Greenland says. "Such a tiny value is saying that the Standard Model without the Higgs boson [can't be correct]. It's screaming at that level."

But physics allows for a level of precision that isn't achievable elsewhere. When you're testing people, as in psychology, you will never achieve odds of one in three million. A *p* value of 0.05 puts the odds of repeated rejection of a correct hypothesis across many tests at one in 20. (It does not indicate, as is often believed, that the chance of error on any single test is 5 percent.) That's why statisticians long ago added "confidence intervals," as a way of providing a sense of the amount of error or uncertainty in estimates made by scientists. Confidence intervals are mathematically related to *p* values. *P* values run from 0 to 1. If you subtract 0.05 from 1, you get 0.95, or 95 percent, the conventional confidence interval. But a confidence interval is simply a useful way of summarizing the results of hypothesis tests for many effect sizes. "There's nothing about them that should inspire any confidence," Greenland says. Yet over time both *p* values and confidence intervals took hold, offering the illusion of certainty.

P values themselves are not necessarily the problem. They are a useful tool when considered in context. That's what journal editors and scientific funders and regulators claim they do. The concern is that the importance of statistical significance might be exaggerated or overemphasized, something that's especially easy to do with small samples. That's what led to the current replication crisis. In 2015 Brian Nosek, co-founder of the Center for Open Science, spearheaded an effort to replicate 100 prominent social

psychology papers, which found that only 36.1 percent could be replicated unambiguously. In 2018 the Social Sciences Replication Project reported on direct replications of 21 experimental studies in the social sciences published in *Nature* and *Science* between 2010 and 2015. They found a significant effect in the same direction as in the original study for 13 (62 percent) of the studies, and the effect size of the replications was on average about half the original effect size.

Genetics also had a replication crisis in the early to mid-2000s. After much debate, the threshold for statistical significance in that field was shifted dramatically. "When you find a new discovery of a genetic variance related to some disease or other phenotype, the standard for statistical significance is 5×10^{-8}, which is basically 0.05 divided by a million," says Benjamin, who has also worked in genetics. "The current generation of human genetics studies is considered very solid."

The same cannot be said for biomedical research, where the risk tends toward false negatives, with researchers reporting no statistical significance when effects exist. The absence of evidence is not evidence of absence, just as the absence of a wedding ring on someone's hand is not proof that the person isn't married, only proof that the person isn't wearing a ring. Such cases sometimes end up in court when corporate liability and consumer safety are at stake.

BLURRING BRIGHT LINES

JUST HOW MUCH TROUBLE is science in? There is fairly wide agreement among scientists in many disciplines that misinterpretation and overemphasis of *p* values and statistical significance are real problems, although some are milder in their diagnosis of its severity than others. "I take the long view," says social psychologist Blair T. Johnson of the University of Connecticut. "Science does this regularly. The pendulum will swing between extremes, and you've got to live with that." The benefit of this round, he says,

© 2020 Scientific American

is that it is a reminder to be modest about inferences. "If we don't have humility as scholars, we're not going to move forward."

To truly move forward, though, scientists must agree on solutions. That is nearly as hard as the practice of statistics itself. "The fear is that taking away this long-established practice of being able to declare things as statistically significant or not would introduce some kind of anarchy to the process," Wasserstein says. Still, suggestions abound. They include changes in statistical methods, in the language used to describe those methods and in the way statistical analyses are used. The most prominent ideas have been put forth in a series of papers that began with the ASA statement in 2016, in which more than two dozen statisticians agreed on several principles for reform. That was followed by a special issue of one of the association's journals that included 45 papers on ways to move beyond statistical significance.

In 2018 a group of 72 scientists published a commentary called "Redefine Statistical Significance" in *Nature Human Behaviour* endorsing a shift in the threshold of statistical significance from 0.05 to 0.005 for claims of new discoveries. (Results between 0.05 and 0.005 would be called "suggestive.") Benjamin, the lead author of that paper, sees this as an imperfect short-term solution but as one that could be implemented immediately. "My worry is that if we don't do something right away, we'll lose the momentum to do the kind of bigger changes that will really improve things, and we'll end up spending all this time arguing over the ideal solution. In the meantime, there will be a lot more damage that gets done." In other words, don't let the perfect be the enemy of the good.

Others say redefining statistical significance does no good at all because the real problem is the very existence of a threshold. In March 2019 U.C.L.A.'s Greenland, Valentin Amrhein, a zoologist at the University of Basel, and Blakeley McShane, a statistician and expert in marketing at Northwestern University, published a comment in *Nature* that argued for abandoning the concept of statistical significance. They suggest p values be used as a continuous variable among other pieces of evidence and confidence intervals be renamed "compatibility intervals" to reflect what they actually signal: compatibility with the data, not confidence in the result. They solicited endorsements for their ideas on Twitter. Eight hundred scientists, including Benjamin, signed on.

Clearly, better—or at least more straightforward—statistical methods are available. Gelman, who frequently criticizes the statistical approaches of others, does not use null hypothesis significance testing in his work at all. He prefers Bayesian methodology, a more direct statistical approach in which one takes initial beliefs, adds in new evidence and updates the beliefs. Greenland is promoting the use of a surprisal, a mathematical quantity that adjusts p values to produce bits (as in computer bits) of information. A p value of 0.05 is only 4.3 bits of information against the null. "That's the equivalent to seeing four heads in a row if someone tosses a coin," Greenland says. "Is that much evidence against the idea that the coin tossing was fair? No. You'll see it occur all the time. That's why 0.05 is such a weak standard." If researchers had to put a surprisal next to every p value, he argues, they would be held to a higher standard. An emphasis on effect sizes, which speak to the magnitude of differences found, would also help.

Improved education about statistics for both scientists and the public could start with making the language of statistics more accessible. Back when Fisher embraced the concept of "significance," the word carried less weight. "It meant 'signifying' but not 'important,'" Greenland says. And it's not surprising that the term "confidence intervals" tends to instill undue, well, confidence.

EMBRACE UNCERTAINTY

STATISTICAL SIGNIFICANCE has fed the human need for certainty. "The original sin is people wanting certainty when it's not appropriate," Gelman says. The time may have come for us to sit with the discomfort of not being sure. If we can do that, the scientific literature will look different. A report about an important finding "should be a paragraph, not a sentence," Wasserstein says. And it shouldn't be based on a single study. Ultimately a successful theory is one that stands up repeatedly to decades of scrutiny.

Small changes are occurring among the powers that be in science. "We agree that p values are sometimes overused or misinterpreted," says Jennifer Zeis, spokesperson for the *New England Journal of Medicine*. "Concluding that a treatment is effective for an outcome if $p < 0.05$ and ineffective if $p > 0.05$ is a reductionist view of medicine and does not always reflect reality." She says the journal's research reports now include fewer p values, and more results are reported with confidence intervals without p values. It is also embracing the principles of open science, such as publishing more detailed research protocols and requiring authors to follow prespecified analysis plans and report when they deviate from them.

At the U.S. Food and Drug Administration, there hasn't been any change to requirements in clinical trials, according to John Scott, director of the division of biostatistics. "I think it's very unlikely that p values will disappear from drug development anytime soon, but I do foresee increasing application of alternative approaches," he says. For instance, there has been greater interest among applicants in using Bayesian inference. "The current debate reflects generally increased awareness of some of the limitations of statistical inference as traditionally practiced."

Johnson, who is the incoming editor at *Psychological Bulletin,* has seen eye to eye with the current editor but says, "I intend to force conformity to fairly stringent standards of reporting. This way I'm sure that everyone knows what happened and why, and they can more easily judge whether methods are valid or have flaws." He also emphasizes the importance of well-executed meta-analyses and systematic reviews as ways of reducing dependence on the results of single studies.

Most critically, a p value "shouldn't be a gatekeeper," McShane says. "Let's take a more holistic and nuanced and evaluative view." That was something that even Ronald Fisher's contemporaries supported. In 1928 two other giants of statistics, Jerzy Neyman and Egon Pearson, wrote of statistical analysis: "The tests themselves give no final verdict but as tools help the worker who is using them to form his final decision." SA

Lydia Denworth is a contributing editor for *Scientific American* and is author of *Friendship: The Evolution, Biology, and Extraordinary Power of Life's Fundamental Bond* (W. W. Norton, 2020).

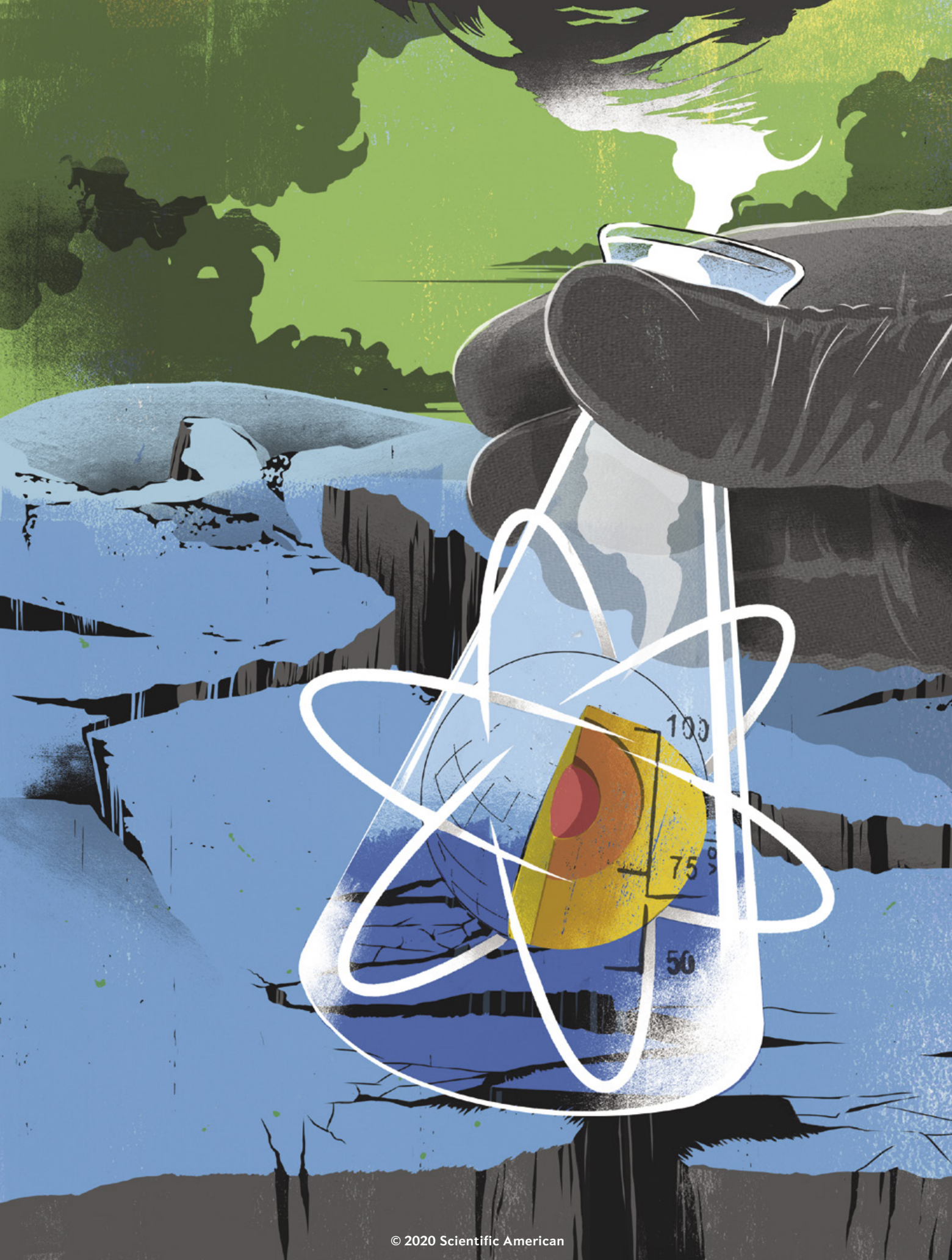
MORE TO EXPLORE

Evaluating the Replicability of Social Science Experiments in *Nature* and *Science* between 2010 and 2015. Colin F. Camerer et al. in *Nature Human Behaviour,* Vol. 2, pages 637–644; September 2018.

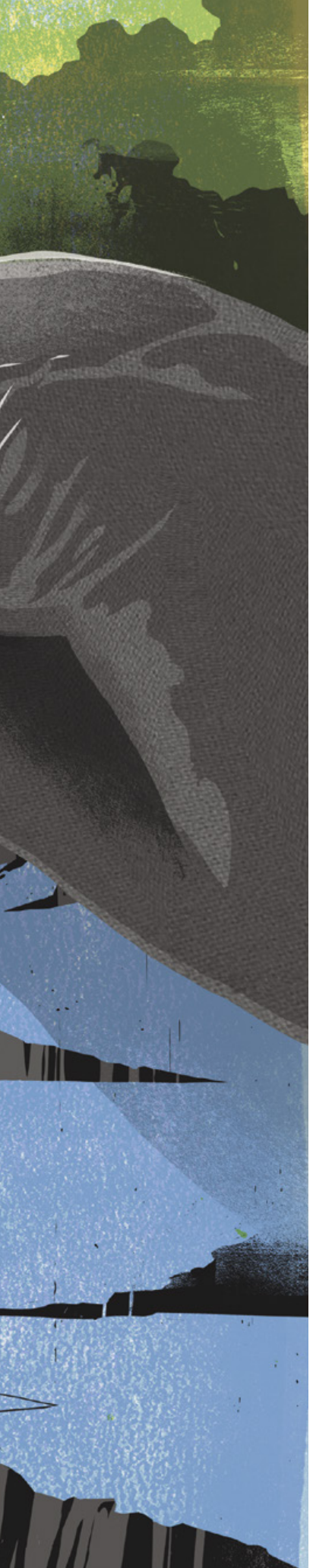
Moving to a World beyond "p< 0.05." Ronald L. Wasserstein, Allen L. Schirm and Nicole A. Lazar in *American Statistician,* Vol. 73, Supplement 1, pages 1–19; 2019.

scientificamerican.com/magazine/sa

© 2020 Scientific American



© 2020 Scientific American



THE ROOTS OF SCIENCE DENIAL

It has nothing do with science itself

By Katharine Hayhoe, as told to Jen Schwartz

Although she's technically Canadian, atmospheric scientist Katharine Hayhoe might understand America's polarized attitudes toward science better than anyone. Her bona fides have serious range: she is co-director of the Climate Science Center and a political science professor at Texas Tech University, CEO of a climate-impact consulting group, creator of the myth-busting Web series Global Weirding *and an electric-car-driving evangelical Christian. Self-described as "on the fringes of many tribes," Hayhoe is equally adept at presenting to church groups and speaking on panels alongside people like Barack Obama and Leonardo DiCaprio. As a result, she has become one of the most lauded and sought-after climate communicators in the country—and the recipient of much hate mail. Hayhoe spoke with* Scientific American *about the war on facts and the forces driving climate skepticism. Edited excerpts from that conversation follow.*

SCIENCE DENIAL IS BASICALLY ANTI-INTELLECTUALISM. IT'S A THREAD THAT has run though American society for decades, possibly even centuries. Back in 1980, Isaac Asimov said that it's "nurtured by the false notion that democracy means that 'my ignorance is just as good as your knowledge.'" Today we're dealing with its most recent manifestation, at its peak.

Climate change is a special case of science denial, which of course goes back to Galileo. The Catholic Church didn't push back on Galileo until he stuck his head out of the ivory tower and published in Italian rather than in Latin, so that he could tell the common people something that was in direct opposition to the church's official program. Same with Darwin. The church didn't have a problem with his theory of evolution until he published a popular book that everyone could read.

IN BRIEF

It's been a grim year. In the U.S. and the U.K, tribalism and anti-intellectualism seem to have triumphed over facts and reason. The effects have rippled around the globe. Some nations, such as China, stand to benefit while historical science powerhouses stumble.

Moving forward will require understanding the cultural and psychological reasons people reject scientific thinking. Many researchers are reconsidering their traditional detachment from politics and learning that public opinion is a force to take seriously.

Illustrations by Jörn Kaspuhl

© 2020 Scientific American

Similarly, we've known about the relationship between carbon dioxide and global warming since the 1890s. It's been more than 50 years since scientists warned President Lyndon B. Johnson about the dangers of a changing climate. But scientists back then didn't get the deluge of hate mail that I get now. So what shifted? It started, possibly, with [Columbia University climate scientist] James Hansen's testimony before Congress in 1988. He announced that a resource we all rely on—and makes many of the world's biggest companies rich—is harming not just the environment but all of humanity. I think it's no accident that Hansen is the most vilified and attacked climate scientist in the U.S. because he was the first person to emerge from the ivory tower and start talking about global warming in a sphere where its implications became apparent for policy and politics.

So you can see that the problem people have with science is never the actual science. People have a problem with the *implications* of science for their worldview and, even more important, for their ideology. When anti-intellectualism rises to the surface, it's because there are new, urgent results coming out of the scientific community that challenge the perspective and status quo of people with power. Renewable energy is now posing a very significant threat to them. The more viable the technologies, the greater the pushback. It's a last-ditch effort to resist change, which is why denial is at a fever pitch.

POLARIZATION IMPLIES TRIBALISM. WE'VE BECOME SO TRIBAL THAT IF YOU'RE ON THE LEFT, IT'S LIKE A STATEMENT OF FAITH TO SAY CLIMATE CHANGE IS REAL. AND IF YOU'RE ON THE RIGHT, IT'S A TENET TO SAY IT ISN'T REAL.

Today, although many of the objections to climate science are couched in science-y terms—*it's just a natural cycle, scientists aren't sure, global cooling, could it be volcanoes*—or even religious-y terms—*God is in control*—99 percent of the time, that language is just a smoke screen. If you refuse to engage these arguments and push through for even five minutes, the conversation will naturally lead to profound objections to climate change *solutions*.

WHAT'S REALLY AT PLAY

THE NUMBER-ONE QUESTION I get from people is, "Could you just talk to my father-in-law, my congressman, my colleague? If you just explain the facts to them, I'm sure it will change their mind." This is a trap. It turns us into Don Quixote, willing to tilt with these people and say, "Here's how we know it's not a natural cycle!" It almost never works. The only way to have a constructive dialogue with a dismissive person is on the level at which he or she really has the issue.

How did the narrative of climate change become a polarized, faith-based system? If we look at surveys, the level of political polarization in the U.S. now compared with 20 or 30 years ago is staggering. Polarization implies a rise in tribalism: an unthinking, unquestioning adherence to the tenets of my tribe. Unfortunately, because climate solutions appear to challenge the ideology of the right-hand side of the political spectrum, it's become one of the most polarized issues in the U.S. We've become so tribal that if you're on the left, it's like a statement of faith to say climate change is real. And if you're on the right, it's a tenet to say climate change isn't real. That's why this "belief" language has come in more naturally rather than artificially.

That said, climate change is deliberately framed as a false religion by those who want people of faith to reject it. You'll see some conservative politicians say, "I'm a true believer, I reject that God is not in charge." It's a very clever messaging technique because if I'm a Christian—and more than 70 percent of Americans are—I'm taught to beware of false prophets. Beware of people saying things that sound good but are actually leading you to worship the created instead of the creator, Earth instead of God.

After presentations to skeptical audiences, I've had people say to me, "You know, this makes sense, and I wish I could agree with you, but I just can't because that would mean I'm agreeing with Al Gore." Any perceived Earth worship immediately triggers an ingrained response to reject. One of the funny images I show in some of my talks is called the Church of Climatology,

© 2020 Scientific American

with Al Gore as the preacher and other politicians and celebrities as the choir. Once somebody photoshopped my head onto one of the choir members. And I thought it was absolutely hilarious because, yes, I get how people feel. We have to laugh together before we can move on to talk about beliefs versus evidence.

That's why Al Gore is one of the best and one of the worst messengers for climate change. The best because he's so passionate and informed and has such a great reach. At the same time—I know he recognizes this—in this politically polarized society, he firmly belongs to only one tribe. So by definition, it means the other tribe must reject him—and everything he stands for.

Climate change, of course, is also a tragedy of the commons, and it requires communal action. Yet the U.S. is the number-one most individualistic country in the world, founded on a revolt against big government and taxes. For many Americans, we have to talk more about market-based and technological solutions that appeal to their values instead of trying to change their identity. Take [Australian cognitive scientist] John Cook, who founded the blog Skeptical Science, which evaluates and pushes back on global warming denial. John couldn't even get his own father to accept climate change. But then his fiscally conservative dad used a rebate program to get solar panels on his house. He saved all kinds of money and started telling everyone how wonderful these panels are. And later, his dad says to John, "You know, this climate change thing, it's probably real, and I'm doing my part." He didn't need to be a wide-eyed tree hugger saving the whales; he could now align climate change with his own identity.

Even in the science community, there's so much confusion over how to communicate. The deficit model—just give them the facts!—does not work in public discourse unless everybody is politically neutral. That's why social science is increasingly important. I was the experimental method in a recent paper where a researcher asked me to speak at an evangelical Christian college. He asked the students about global warming before and after my talk and found statistically significant differences on their perspectives. Many people are now doing this kind of message testing. How humans interact with information is an emerging area of research that's desperately important.

Scientists also tend to understate the impact of climate change. We tend to, in the words of one researcher, "[err] on the side of least drama." We tracked 20 years' worth of studies and found that we systematically underestimate the rate and speed of change. Climate science is under such a microscope now that we like to be 99.9 percent sure of results before we say anything. But are we being too conservative? It's a challenge I confront every day.

THE WORK AHEAD

LOOK, WE CAN'T FIX ALL these issues—cultural, political, psychological—before we take necessary action on climate change. People say to me, "Well, if you could just get everyone onboard with the science…." I'm like, good luck with that! How did that work out the past few centuries? This climate problem is urgent. The window is closing. We have to fix it with the flawed, imperfect society we have today.

We have to start by asking what people's values are, where they're coming from, what they love, what they fear, what gets them up in the morning. I say, "We can agree to disagree, but don't you support solar energy bringing all these jobs to Texas? Did you know Fort Hood gets energy from solar because it's cheaper?" If someone thinks solar power protects us from immigrants or terrorists or the Antichrist, then great, fine. With some groups, I don't even use the words "climate" and "change" sequentially. With Christians, we talk about the Bible's message of stewardship. With libertarians, we talk about free-market strategies. With moms' groups, we talk about pollution affecting our kids' health. With farmers, I say, "Hey, you're the backbone of our food system, how have drought patterns changed?" I don't validate the concept that there is a left and right side to climate science. And neither should the media. We should focus instead on solutions and impacts.

My number-one piece of advice for people doing climate—or any science—outreach is, Don't focus on the dismissive people. They're really a very small part of the population, and they're primarily older white men. Granted, the majority of them seem to be clustered in Washington, D.C., these days. Still, for people who react so emotionally, it's because they've staked their identity on that denial. It's as much a part of them as their kidneys or heart. When you're asking them to change their mind, you are literally seen as a threat. It's worth standing up to them in a public forum and saying, "You are lost. Here is the evidence." Not for the purpose of changing their mind but to show everybody else that we have answers.

Because here's the thing: If you look at Yale University's climate communication surveys, most Americans agree that climate change is real, that humans are causing it and that it's important to do something about it. But the number-one problem we're facing is that most Americans do not think climate change affects them personally. They think it's a problem for poor people in poor counties or for future generations. It's in our psychology to deny an overwhelming problem that isn't immediately bearing down on us. And until recently, we've been shielded by our infrastructure, our crop insurance and home insurance programs. Of course, all of that is up against the wall now, and it's my job to connect those dots.

That's why we [the authors of the government's National Climate Assessment] decided to write a supplemental Climate Science Special Report in 2017. It's the first time we've done it, and it's the most comprehensive, definitive report on climate change that the government has ever published. We made a lot of effort to write in a language that people can understand, and I think it really shuts down the whole "blue versus red" debate. It brings the science down to the level of where we live. You can see how climate change is affecting our food, water, economy, agriculture, infrastructure and security.

The goal of the report is to provide a scientific basis for anyone who wants to know both broadly and specifically why climate change matters to us, now. Many, many more people in this country are in the cautious, disengaged category, but they often seem very quiet. We have to filter out the noise from the dismissive people and talk with those who are lurking at the edges, listening, not sure what they think yet about what should be done but open to dialogue. So forget this elaborate smoke screen. By falling for the illusion that climate deniers can be convinced with more facts, we are distracted from engaging with a much larger group of people who want to understand why and how we should move forward with solutions. And that's exactly what the deniers want. SA

Jen Schwartz is senior editor for technology and mind for *Scientific American*.

© 2020 Scientific American

Message Control

Scientists are trying new ways to win over a skeptical public

By Brooke Borel

IN 1999 ROBERT SHAPIRO, THEN HEAD OF MONSANTO, gave a stunning mea culpa at a Greenpeace conference in London. Monsanto's first genetically engineered (GE) crops had been on the market for only three years, but they were facing fierce public backlash. After a botched rollout marred by lack of transparency, the company, Shapiro said, had responded with debate instead of dialogue. "Our confidence in this technology … has widely been seen, and understandably so, as condescension or indeed arrogance," he said. "Because we thought it was our job to persuade, too often we've forgotten to listen."

The damage was already done. Fifteen years later only 37 percent of the U.S. population thought that GE foods were safe to eat, compared with 88 percent of scientists, according to the Pew Research Center. Regulatory bodies in the U.S. fought for years over whether and how to label GE foods. In 2015 more than half of the European Union banned the crops entirely.

Science doesn't happen in a vacuum. But historically, many researchers haven't done a great job of confronting—or even acknowledging—the entangled relation between their work and how it is perceived once it leaves the lab. "The dismal experience we had with genetically engineered foods was an object lesson in what happens when there's a failure to engage the public with accurate information and give them an opportunity to think through trade-offs for themselves," says R. Alta Charo, a bioethicist and professor of law at the University of Wisconsin–Madison. When communication breaks down between science and the society it serves, the resulting confusion and distrust muddies everything from research to industry investment to regulation.

In the emerging era of CRISPR and gene drives, scientists don't want to repeat the same mistakes. These new tools give researchers an unprecedented ability to edit the DNA of any living thing—and, in the case of gene drives, to alter the DNA of wild populations. The breakthroughs could address big global problems, curtailing health menaces such as malaria and breeding crops that better withstand climate change. Even if the expectations of CRISPR and gene drives do come to frui-

Illustration by Jörn Kaspuhl

tion—and relevant products are safe for both people and the environment—what good is the most promising technology if the public rejects it?

"Without transparency, we might see a kind of hyperpolarization," says Jason Delborne, a professor of science, policy and society at North Carolina State University. Concerned groups will feel marginalized, and advocates won't receive critical feedback needed to improve design and safety. "This puts the technology at risk of a knee-jerk moratorium at the first sign of difficulty," he notes.

To avoid that outcome, some researchers are taking a new tack. Rather than dropping fully formed technology on the public, they are proactively seeking comments and reactions, sometimes before research even starts. That doesn't mean political and social conflict will go away entirely, Delborne says, "but it does contribute to a context for more democratic innovation." By opening an early dialogue with regulators, environmental groups and communities where the tools may be deployed, scientists are actually tweaking their research plans while wresting more control over the narrative of their work.

Take evolutionary geneticist Austin Burt. In 2003 he published the first theoretical paper on GE gene drives. Shortly after, with funding from the Bill & Melinda Gates Foundation, he and his colleagues launched a research project to see if gene drives could control *Anopheles* mosquitoes, which spread malaria. Back then, in the pre-CRISPR days, the technology was so speculative that doing outreach "didn't seem worth taking up people's time," Burt says. Now that a working gene drive may be ready for regulatory assessment within five years, it's essential to talk to communities where the technology may be deployed, he adds, "so we can make things that are going to be acceptable not just to regulators but to the public at large."

This push for reflection has especially come from those wielding the checkbooks. In 2016 the National Academies of Sciences, Engineering, and Medicine published *Gene Drives on the Horizon: Advancing Science, Navigating Uncertainty, and Aligning Research with Public Values.* The sponsors—various federal agencies, the Gates Foundation and the Foundation for the National Institutes of Health—specifically asked for comprehensive recommendations on ethics and public engagement, says Keegan Sawyer, project director for the report. Other National Academies reports have included these elements, but the combination that appeared in the gene drive report was "unusual," Sawyer says.

DARPA is among those listening to the guidelines. Its Safe Genes initiative, which will fund seven research projects aimed at understanding how to deploy and control gene drives, requires all its projects to have thorough public engagement plans. One DARPA grant recipient is a team at N.C. State, which includes Delborne. He is overseeing social engagement on a gene drive project that aims to remove invasive mice from remote islands to protect seabirds and other wildlife. Although the research is underway, Delborne says the partners "have been really clear since the very beginning that if people reject this technology for ethical reasons or because there are concerns about the risks—even if the scientists don't see it that way—there is essentially a pathway to no." Simply put, the scientists are willing to halt the project.

On the even more extreme end of this trend is Kevin Esvelt, an evolutionary engineer at the Massachusetts Institute of Technology. He's considering genetic technologies to alter wild mice so that they cannot carry and spread the pathogen that causes Lyme disease. In 2016, before starting any work in the lab, Esvelt visited Lyme-plagued Nantucket, Mass., to gauge if residents would be interested in genetic approaches—including gene drives, although he advised against this option because he doesn't think it is suitable in this case. Nantucket followed Esvelt's lead on gene drives, although the community is exploring the possibility of an alternative technology to immunize mice against the pathogen.

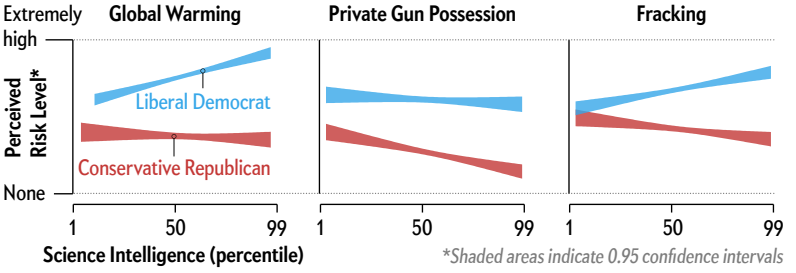
Esvelt was addressing head-on a special ethical quandary of gene drives, which are designed to spread and persist in the shared environment: Who should get to decide whether and how to use such technology? "To me, it is mind-boggling that we got so much attention just for going to the communities before we did anything else," Esvelt says. "I think that says something about how science is typically done."

Whether the emergence of these efforts will reduce fear and skepticism "depends on how responsive the people listening to the engagement are to those concerns," says Jennifer Kuzma, who is co-director of the Genetic Engineering and Society Center at N.C. State. In other words, researchers must be willing not only to hear the public's confusion and pushback but also to adapt—even if that means shelving a technology they think could change the world. SA

Brooke Borel is a journalist and author who frequently reports on biotechnology.

The Deficient Deficit Model

The gold standard for science communication used to be the deficit model, which assumed that the key to acceptance was education. Just explain the science of vaccines or GMOs to the skeptical public, and distrust will fall away. But recent work by researchers, including Dan M. Kahan of Yale Law School, suggests a more complicated relation that involves personal identity and belief. Kahan's work shows that people with contrasting political values will draw different conclusions from the same evidence even when they are scientifically literate.



SOURCE: "WHAT IS THE 'SCIENCE OF SCIENCE COMMUNICATION'?" BY DAN M. KAHAN, IN *JOURNAL OF SCIENCE COMMUNICATION*, VOL. 14, NO. 3. PUBLISHED ONLINE AUGUST 25, 2015

Graphic by Amanda Montañez

How Much Can We Know?

The reach of the scientific method is constrained by the limitations of our tools and the intrinsic impenetrability of some of nature's deepest questions

By Marcelo Gleiser

"WHAT WE OBSERVE IS NOT NATURE IN ITSELF but nature exposed to our method of questioning," wrote German physicist Werner Heisenberg, who was the first to fathom the uncertainty inherent in quantum physics. To those who think of science as a direct path to the truth about the world, this quote must be surprising, perhaps even upsetting. Is Heisenberg saying that our scientific theories are contingent on us as observers? If he is, and we take him seriously, does this mean that what we call scientific truth is nothing but a big illusion?

People will quickly counterstrike with something like: Why do airplanes fly or antibiotics work? Why are we able to build machines that process information with such amazing efficiency? Surely, such inventions and so many others are based on laws of nature that function independently of us. There is order in the universe, and science gradually uncovers this order.

No question about it: There is order in the universe, and much of science is about finding patterns of behavior—from quarks to mammals to galaxies—that we translate into general laws. We strip away unnecessary complications and focus on what is essential, the core properties of the system we are studying. We then build a descriptive narrative of how the system behaves, which, in the best cases, is also predictive.

Often overlooked in the excitement of research is that the methodology of science requires interaction with the system we are studying. We observe its behavior, measure its properties, and build mathematical or conceptual models to understand it better. And to do this, we need tools that extend into realms beyond our sensorial reach: the very small, the very fast, the very distant and the virtually inaccessible, such as what is inside the brain or buried in the earth's core. What we observe is not nature itself but nature as discerned through data we collect from machines. In consequence, the scientific worldview depends on the information we can acquire through our instruments. And given that our tools are limited, our view of the world is necessarily myopic. We can see only so far into the nature of things, and our ever shifting scientific worldview reflects this fundamental limitation on how we perceive reality.

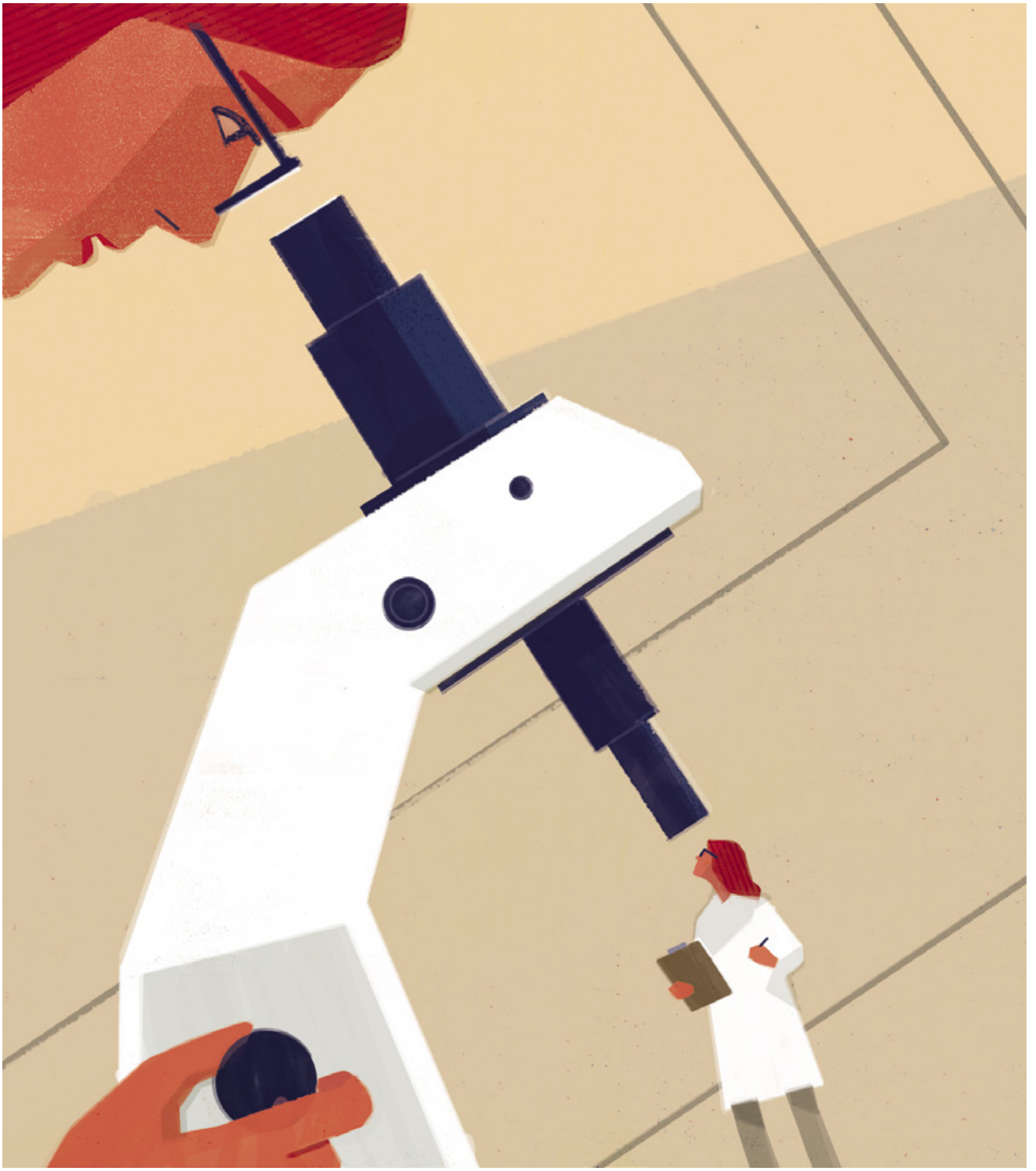
Just think of biology before and after the microscope or gene sequencing, or of astronomy before and after the telescope, or of particle physics before and after colliders or fast electronics. Now, as in the 17th century, the theories we build and the worldviews we construct change as our tools of exploration transform. This trend is the trademark of science.

Sometimes people take this statement about the limitation of scientific knowledge as being defeatist: "If we can't get to the bottom of things, why bother?" This kind of response is misplaced. There is nothing defeatist in understanding the limitations of the scientific approach to knowledge. Science remains our best methodology to build consensus about the workings of nature. What should change is a sense of scientific triumphalism—the belief that no question is beyond the reach of scientific discourse.

There are clear unknowables in science—reasonable questions that, unless currently accepted laws of nature are violated, we cannot find answers to. One example is the multiverse: the conjecture that our universe is but one among a multitude of others, each potentially with a different set of laws of nature. Other universes lie outside our causal horizon, meaning that we cannot receive or send signals to them. Any evidence for their existence would be circumstantial: for example, scars in the radiation permeating space because of a past collision with a neighboring universe.

Other examples of unknowables can be conflated into three questions about origins: of the universe, of life and of the mind. Scientific accounts of the origin of the universe are incomplete because they must rely on a conceptual framework to even begin to work: energy conservation, relativity, quantum physics, for instance. Why does the universe operate under these laws and not others?

© 2020 Scientific American



Similarly, unless we can prove that only one or very few biochemical pathways exist from nonlife to life, we cannot know for sure how life originated on Earth. For consciousness, the problem is the jump from the material to the subjective—for example, from firing neurons to the experience of pain or the color red. Perhaps some kind of rudimentary consciousness could emerge in a sufficiently complex machine. But how could we tell? How do we establish—as opposed to conjecture—that something is conscious?

Paradoxically, it is through our consciousness that we make sense of the world, even if only imperfectly. Can we fully understand something of which we are a part? Like the mythic snake that bites its own tail, we are stuck within a circle that begins and ends with our lived experience of the world. We cannot detach our descriptions of reality from how we experience reality. This is the playing field where the game of science unfolds, and if we play by the rules we can see only so much of what lies beyond. SA

Marcelo Gleiser is Appleton Professor of Natural Philosophy and a professor of physics and astronomy at Dartmouth College. He has authored several books, including *The Island of Knowledge: The Limits of Science and the Search for Meaning* (2014). In 2019 he was awarded the Templeton Prize.

Illustration by Jessica Fortner

© 2020 Scientific American

Ya Know?

Beyond the unknown unknowns is what's unknowable

By Steve Mirsky

In the 1954 World Series, Willie Mays of the New York Giants made what many consider the greatest catch in baseball history on a long fly ball to straightaway centerfield hit by Vic Wertz of the Cleveland Indians. Broadcaster Bob Costas talked about the catch for the Ken Burns documentary series *Baseball:* "It was more than just a great catch. It was a catch no one had ever seen before … it was a play that until that point was outside the realm of possibility in baseball." Mays in that moment thus expanded baseball into previously nonexistent territory, much like the universe expands—and not *into* anything, for there was nothing there before.

On the other hand, nah. Indians' pitcher Bob Feller, who watched the play from the dugout, followed Costas on the episode. "It was far from the best catch I've ever seen," he says. "It was a very good catch. We knew Willie had the ball all the way."

I thought of this sequence more than once when I attended a conference in 2019 at the New School's Center for Public Scholarship here in New York City billed as "Unknowability: How Do We Know What Cannot Be Known?" Filled with doubt, I felt fortunate to simply find the auditorium.

Discussing the unknown, Columbia University biologist Stuart Firestein cited what he called an apocryphal saying: "It's very hard to find a black cat in a dark room, especially when there is no cat." He continued, "I think this is exactly how science works and how it deals with the so-called unknowable. We stomp around in black rooms and eventually … we may find this critter or we may find some other critter entirely. But once having decided the room is either empty or full of a cat, we simply move on to the next dark room."

He also cited James Clerk Maxwell as having said, "'Thoroughly conscious ignorance is the prelude to every real advance in science.'" Firestein went on, "And so this is the kind of ignorance that I'm talking about, not the common usage of the word 'ignorance,' not stupidity or willful indifference to fact or logic—you know who I'm talking about. But rather this thoroughly conscious kind of ignorance that can be developed … The big question for me really is we've gained some knowledge, what does one do with that knowledge? And the purpose of that knowledge in my opinion is to create better ignorance, if you will. Because there's low-quality ignorance and high-quality ignorance … science, in my opinion, is the search for better ignorance." Presumably, as the quality of ignorance increases, so does the level of associated bliss.

After University of Cambridge mathematician John Barrow pointed out that "the unknown … is of course a vast, untapped field—rather like studying everything that is not a banana," he mentioned that beyond unknown unknowns lies the truly unknowable. "[Kurt] Gödel announced that … if you have a system that's got a finite number of axioms … and if it's complicated enough to include arithmetic … and if it's consistent … then there are statements of arithmetic which you can neither show to be true nor false using the rules and axioms of arithmetic."

Gödel's knack for deep insights led to a famous story about his U.S. citizenship interview. He allegedly cheerfully announced that he had discovered a way to apply the Constitution that would turn the U.S. into a dictatorship. (See above, "You know who I'm talking about.") Legend has it that his friend Einstein, on hand for the happy day, jumped in to change the subject.

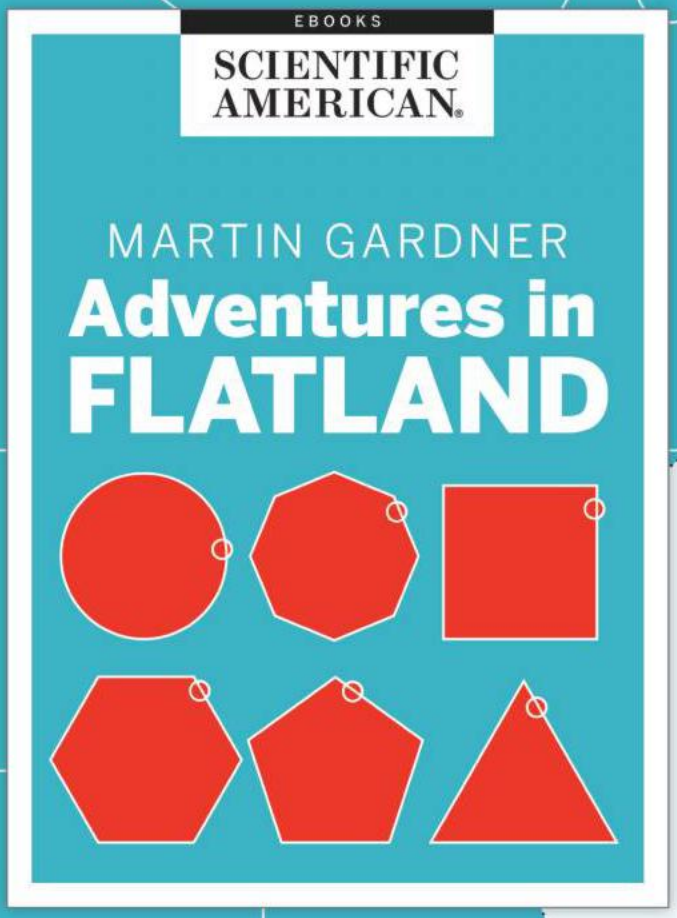
Uncertainty and unknowability may feel discouraging. But Firestein thought they could be a source of optimism, as in the story of the condemned prisoner who convinces the king to give him a year's reprieve in return for the promise that the inmate will teach the monarch's horse to talk.

Another prisoner asks the saved man what possessed him to make such a crazy bargain. "The fellow says, 'A lot can happen in a year. The horse might die. The king might die. I might die. The horse might learn to talk.'" That last option may seem overly optimistic. But it certainly beats the alternative. SA

Steve Mirsky has been writing the Anti Gravity column since a typical tectonic plate was about 36 inches from its current location. He also hosts the *Scientific American* podcast Science Talk.

Illustration by Matt Collins

© 2020 Scientific American



Playing Games in 2D

Celebrated mathematician Martin Gardner mixed well-understood topics with the cutting-edge. In this eBook we've selected his best columns set in two dimensions, including geometric puzzles of transforming one shape to another, the physics of a 2D world and more.

sciam.com/ebooks

SCIENTIFIC AMERICAN eBooks

Scientific American is a registered trademark of Springer Nature America, Inc.

Expertise. Insights. Illumination.

Discover world-changing science. Get 12 issues of *Scientific American* in print and explore our digital archive back to 1845, including articles by more than 150 Nobel Prize winners.

sciam.com/print&archive

Download on the App Store

GET IT ON Google Play

SPECIAL ISSUE : SEX AND GENDER
GENDER MYTHS DEBUNKED
UNHEALTH PRACTICE
SPEAKING IN WHISTLES
WHAT'S WRONG WITH COSMIC INFLATION?
SCIENTIFIC AMERICAN
WORKOUTS AND WEIGHT LOSS
Learn the surprising evolutionary reason why exercise alon won't shed pounds—and what to do about it
PLUS
PREVENTING THREATS TO CROPS
Airborne microbes' critical role PAGE 40
WHAT IS A KILOGRAM?
Science seeks new ways to ensure accuracy PAGE 46
MEDICAL SCANS
A shortage of rare atoms imperils disease imaging PAGE 68
SCIENTIFIC AMERICAN
AUTOMOBILE NUME
SCIENTIFIC AMERICAN
T. REX reexamined
SCIENTIFIC AMERICAN
September 1999

Scientific American is a registered trademark of Springer Nature America, Inc. Google Play and the Google Play logo are trademarks of Google LLC. Apple, the Apple logo, iPhone, and iPad are trademarks of Apple Inc., registered in the U.S. and other countries and regions. App Store is a service mark of Apple Inc.